

AIDS at Christmas time

The end of 2005 was supposed to mark the achievement of a critical goal in the treatment of HIV in poor countries. The goal hasn't been met, but it is now within sight.

Two years ago, the United Nations and the World Health Organization (WHO) launched the 3 By 5 Initiative for the global treatment of HIV, with the aim of providing 3 million people in developing countries with antiretroviral drugs by the end of 2005. This ambitious target will not quite be met, with the number falling short by at least a million. But there has been a great advance on the 400,000 who were receiving treatment at the end of 2003. That momentum must be sustained into the new year.

Continued progress will depend on strong political leadership in the countries hardest hit by AIDS, as well as on cash support from outside. The issue of drug pricing has become less acute, as mechanisms have been established to supply HIV treatments at a reasonable cost. But more than 4 million additional patients need the drugs now and tens of millions more will eventually require them.

The initiative set specific national targets, and these have already been met in some middle-income countries; poorer countries are having more trouble, on account of the chronic weakness of their public-health systems. But even here there are grounds for hope. Malawi, for example, has increased the number of its people who are receiving antiretroviral treatment from just 4,000 in 2003 to about 36,000. With more cash support, its programmes can expand to reach the estimated 100,000 other Malawians who still need antiretroviral therapy.

Elsewhere, leaders have overcome cultural barriers and the stigma of AIDS: the prime minister of Lesotho, for example, was tested for HIV in public. And since Brazil initiated free treatment in 1996, deaths from AIDS-related hospitalizations have declined by four-fifths. Yet in too many regions of the world, drug availability remains chronically inadequate. Against the successes of nations such as Malawi and Brazil must be set the failures of others, including three countries with some of the biggest AIDS crises of all: South Africa, Nigeria and India.

South Africa is one of the wealthiest countries on the African continent, but less than a fifth of the nearly 700,000 people who need drugs are receiving them. Manto Tshabalala-Msimang, the country's

health minister, meanwhile continues to emphasize herbal remedies, most recently in a speech in Durban on 1 December.

In Nigeria, an inept and corrupt bureaucracy has severely impeded the roll-out of treatment, which remains out of reach for the overwhelming bulk of the estimated 500,000 people who need it. The government has failed to substantially boost health spending, despite recent windfall revenues from oil exports.

And in India, which may be on the brink of an explosive HIV epidemic, access to treatment has been slow to improve and government officials have been reluctant to face up to the likely extent of the problem. Prime Minister Manmohan Singh has at least shown some leadership in this regard, calling earlier this month for people to shed traditional inhibitions about discussing sex and to address the threat head on.

Next spring, the WHO will set revised targets for access to the medicines, as it moves towards its existing goal of 'universal access' to appropriate therapies by 2010 — a goal endorsed by world leaders last July at the G8 meeting at Gleneagles in Scotland. But for that to happen, the world's richest nations need to provide money, particularly for the Global Fund to Fight AIDS, Tuberculosis and Malaria, which has so far raised only \$3.7 billion of the \$7 billion that it would like to spend by 2007. Major corporations should also contribute directly to the Global Fund — an approach endorsed recently by the Global Business Coalition on HIV/AIDS, whose members include British American Tobacco and Anglo American.

Only a few years ago, antiretrovirals cost thousands of dollars per patient and widespread doubts persisted about their efficacy in places that lacked a good public-health infrastructure. The goal of universal access seemed wildly remote. This Christmas, it seems much closer. Effective HIV treatment can be widely introduced and administered, even in the poorest countries. The world must move forward rapidly towards universal access. ■

"Effective HIV treatment can be widely introduced and administered, even in the poorest countries."

A poor assessment

Given Japan's strong scientific record, the country has a badly flawed research evaluation system.

In the next few weeks, the government of Japan will announce its budget for the fiscal year starting in April 2006. The slow economy and tight overall budget situation may finally have caught up with research, and this year, for the first time in fifteen years, science spending could be cut.

Economies and budgets wax and wane, and scientists cannot expect increased funds as a birthright. But they do have a right to expect fair and transparent evaluation as a guide to good budget management. Japan's national system is letting them down. For decades after the Second World War, spending on science was distributed evenly among about a hundred national universities. But since the mid-1990s, Japan has taken a more selective approach, as befits one of the world's leading scientific powers.

The Council for Science and Technology Policy was established in 2001 to advise the prime minister. Its 15-member council, chaired by the prime minister and including five other ministers of state,



industry representatives and a few scientists, carries out an annual evaluation of every science project funded by government agencies. It uses subcommittees to prioritize according to four grades — S (for superior), A, B and C — on the basis of scientific innovation, international competitiveness and degree of social contribution.

Increasingly, and this year in particular by all accounts, the system bears little resemblance to an objective, independent assessment. This can be a serious problem for major initiatives involving numerous laboratories and hundreds of millions of yen. One problem is a quota system for grades that can be arbitrary and unfair. Such grade quotas need not be a problem if they are applied on a sliding scale that takes into account objective, well-based judgements of achievement across disciplines. But that is not what happens. Too often, judgements, often based on a single day's visit to a project's group leader, don't do more than scratch the surface of a project's significance.

Another problem is that the committee is entirely Japanese. There is of course a limit to how much international experts can be involved. But an international perspective would seem obligatory, particularly when assessing large projects, some of which depend on international collaboration and represent a world-class effort costing many billions of yen.

But the worst failing of the system is a progressive distortion of supposedly objective assessment by the priorities and preferences of

the committee and government. After discussions in closed rooms, ratings emerge that in many cases bear no relationship to scientific achievement or potential, and seem to defy explanation. A major project may be graded 'S' for two years in a row and then be graded A despite maintaining its performance. Even worse, some cutting-edge projects, after many years of top-level grades, have this year been graded 'C' for no conceivable scientific reason.

Some might argue that scientific spending, like other funding, must follow government priorities and so be subject to abrupt changes. No one would suggest that national priorities should remain fixed. But for the process to be nationally and internationally credible, and for top-notch scientists to believe that Japan is a good place to spend their best years, the system of evaluation must be revised. Many researchers see it as opaque and apparently arbitrary. Japan may not be unique — other leading countries also lack a clear evaluation process — but this does not make it acceptable.

Scientific assessment should be objective, well considered and transparent to those being assessed. It should be kept distinct from the process of priority setting, which should itself be open, and should involve greater participation of researchers before final decisions are reached. ■

"Scientific assessment should be objective, well considered and transparent to those being assessed."

A recipe for trouble

A prestigious research agency should have thought twice before attaching its name to a diet book.

Going on a diet is a popular new year's resolution. This year, a diet book penned by researchers in Australia is set to turn up in many Christmas stockings. But its runaway success could damage the reputation of Australia's foremost research institution (see page 1060).

The diet book in question is by no means ground-breaking. Its high-protein message is not that different from others that have drifted into fashion in the past few years. But this one bears the badge of Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO).

In some parts of the world, it might seem odd that splashing the name of a scientific institution on the cover would shift copies of a book to the public. But CSIRO — which runs Australia's main network of government laboratories — has an unusually good public reputation. It is widely perceived as a trusted national institution. Its history, including its pivotal role in the development of agriculture and mining in Australia, has left a strong impression that it knows how to put science to good use.

But the commercial success of the book, which knocked Harry Potter and *The Da Vinci Code* off the national bestseller perch, is irritating some scientists, and for good reason.

The benefits of a high-protein diet remain a hot topic of debate among nutritionists. But even some of those who approve of such a diet question whether it should rely as heavily on meat as this one

does, given the health risks associated with high meat consumption.

But what really rankles with the book's critics is the way it is being marketed. There's something decidedly unsavoury about using the phrase "scientifically proven" to sell anything to a trusting public, yet this is writ large on the book's front cover. The diet is also being promoted as being beneficial for everyone, whereas the published research indicates that it is superior to a high-carbohydrate diet only for a subpopulation of overweight women with symptoms of metabolic dysfunction.

Furthermore, the research behind the book was largely funded by the meat and dairy industries, whose products feature prominently in the diet. Detractors say that this aspect should have been more explicitly recognized, instead of being buried in the book's acknowledgements. The authors insist that the sponsors had no influence on the book's content, but the impression remains of a conflict of interest.

To be fair, the book was not the idea of the researchers or even CSIRO's management. It came from a wily commercial publisher who spotted an opportunity. CSIRO, which has its own publishing arm, only reaps a small percentage of the profits in the form of royalties to its nutritional-research division.

Defenders of the book will argue that its success illustrates how to translate research into an accessible and popular format that puts science into practice. But that argument doesn't justify CSIRO giving permission for its name to be used in a way that could ultimately taint its hard-earned reputation. ■

"There's something decidedly unsavoury about using the phrase 'scientifically proven' to sell anything to the public, yet this is writ large on the book's front cover."

RESEARCH HIGHLIGHTS



Skin deep

Science **310**, 1782–1786 (2005)

Geneticists say they have identified a gene that plays a role in determining differences in people's skin colour.

Previous studies have identified many gene changes responsible for rare disorders such as albinism. But now, Keith Cheng at the

Pennsylvania State University College of Medicine in Hershey and his colleagues have found a gene variation that explains 25–38% of the difference in skin colour between populations of European and African ancestry. The key clue came from zebrafish (*Danio*

rerio) whose skin colour was lighter than normal. Such fish had a mutation causing a shortening of the pigmentation protein Slc24a5. Darker fish had a longer version of the same protein. A single mutation in the human *SLC24A5* gene was found to be shared by people with light skin.

PHOTONICS

Photons learn to crawl

Phys. Rev. Lett. **95**, 253601 (2005)

Slowing down light has become a party trick for physicists. They have previously brought it to a virtual standstill by exploiting exotic optical properties in various media, turning them into 'optical molasses'. An ordinary ruby crystal, for example, will slow light down to about 50 metres per second.

Now Pengfei Wu and Devulapalli Rao of the University of Massachusetts in Boston have shown that the optical properties of the bacterial light-absorbing pigment bacteriorhodopsin, embedded in a thin polymer film, can be manipulated with a laser beam to slow pulses in a second beam to just 0.09 millimetres per second.

Cheap and versatile, these films could have applications in optical technology, such as switching.

BIOTECHNOLOGY

Bumper crop

Nature Biotechnol. doi:10.1038/nbt1173 (2005)

Scientists trying to breed high-yielding crops often find that traits that boost growth are offset by side effects that compromise the plant's yield. A team in Japan reports a way around this problem in rice, and says the findings could be adapted to improve other cereal crops, reducing the need for artificial fertilizers.

Tomoaki Sakamoto of the University of Tokyo and his colleagues screened rice plants for mutations that made the leaves grow at a more erect angle. This should reduce the shade cast on the lower leaves of the plant, increasing photosynthesis. One such

mutation improved grain yield when the rice was densely planted, without weakening the plants. The affected gene produces a plant hormone that could be targeted in other crops.

CHEMISTRY

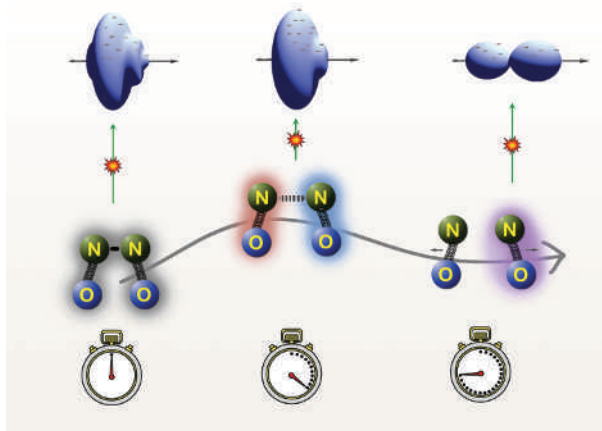
Quick as a flash

Science doi:10.1126/science.1120779 (2005)

Short flashes of laser light lasting mere femtoseconds (10^{-15} seconds) can reveal the progress of chemical reactions by providing snapshots of changing molecules. But the random orientations of these molecules tend to cloud the picture.

Now a team led by Albert Stolow at the National Research Council Canada in Ottawa has used spectroscopic techniques to watch the dissociation of nitric oxide dimers from an individual molecule's point of view, thus removing the randomizing effect.

The observations reveal previously obscure details about how the electron cloud around the molecule changes shape (pictured) during the transformation, offering potential for 'filming' the details of chemical processes.



Firing ultra-short pulses of laser light at nitrogen molecules shows how the electron cloud (top) changes as the two atoms pull apart.

CANCER BIOLOGY

Divide on regardless

Cancer Cell **8**, 479–484 (2005)

Immature cells and stem cells are more likely than specialized cells to go ahead and divide if their chromosomes are entangled, making them naturally more prone to becoming cancerous, say researchers.

Timothy Bestor of Columbia University in New York and his colleagues studied multipotent progenitor cells — which divide to give rise to more specialized cells — in humans and mice. They used a drug to block the action of an enzyme that normally disentangles chromosomes, and found that a greater proportion of progenitor cells pressed on with cell division, compared with similarly treated specialized cells.

Cells that divide regardless produce daughters with damaged and abnormal chromosomes. The research team suggests that such aberrations could help to give rise to cancer stem cells, a subset of tumour cells thought to be the driving force in cancer development.

MICROBIOLOGY

Romping *Rickettsia*

Cell **123**, 1013–1023 (2005)

The bacterium that causes life-threatening Mediterranean spotted fever uses a molecular stooge to get inside cells, say researchers in France. *Rickettsia conorii* is transmitted by ticks and has been classified as a possible bioterrorism agent. But how it penetrates mammalian host cells has long been a mystery.

Pascale Cossart from the Pasteur Institute in Paris and her colleagues have now identified two key

proteins involved in the process — mammalian Ku70 and bacterial rOmpB. Ku70 is normally found in the nucleus of mammalian cells. But it can also move to the cell membrane, where *Rickettsia*'s rOmpB protein can grab it and use it to invade the cell.

STEM CELLS

Wired for action

Proc. Natl Acad. Sci. USA **102**, 18638–18643 (2005)

Human embryonic stem cells can form functional adult neurons when implanted into mouse embryo brains. It has been shown before that such cells can form neurons *in vitro* (pictured above) or when transplanted, but not that they could respond to signals within the mouse brain to form electrically connected neurons.

An international team headed by Fred Gage at The Salk Institute for Biological Studies in La Jolla, California, found that, when injected into the mouse brain, human embryonic stem cells specialized to form neurons and supporting glial cells.

The researchers performed electrical recordings on brain slices from these mice and found that the human neurons had similar electrical properties to regular mouse neurons, and transmitted electrical impulses when stimulated.

MEDICINE

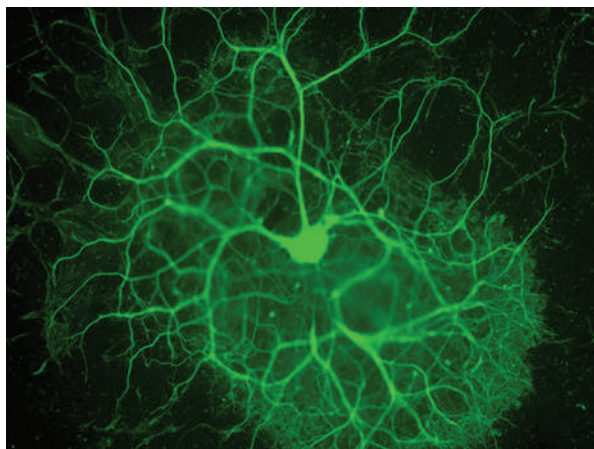
Gut feeling

J. Exp. Med. **202**, 1703–1713 (2005)

Carbon monoxide is toxic at high concentrations, but inhaling small quantities of this gas may give relief to patients with diseases of the gut, say researchers.

Cigarette smoke has been known for decades to protect against chronic ulcerative colitis, an inflammation of the gut triggered by intestinal microbes. Now Scott Plevy and his colleagues at the University of Pittsburgh, Pennsylvania, and Harvard Medical School in Boston, Massachusetts, report that the carbon monoxide in the smoke could account for the protective effect.

In mouse studies, they found that inhaling the gas inhibits production of the immune-cell protein IL-12, which drives gut inflammation. Although carbon monoxide is already known to inhibit acute inflammation, this study is the first to show that it can inhibit established chronic inflammation.



PROC. NATL. ACAD. SCI. USA

CELL BIOLOGY

A light stretch or bend

Nature Chem. Bio. doi:10.1038/nchembio756 (2005)

Protein channels in the cell membranes of neurons open and close in response to chemical neurotransmitters. Now cell biologists have developed a tool that allows exceptional control over this process in channels that respond to the important neurotransmitter glutamate.

Dirk Trauner of the University of California, Berkeley, and his colleagues worked with a compound known to make the channel open. They tethered it to the channel with a molecule that bends or stretches out, depending on the wavelength of light shone on it.

The team showed that bending the linker molecule brings the activating compound closer to the channel, causing it to open. Others may be able to use these new 'light sensitive' channels to re-engineer and explore neurological pathways.

MICROFLUIDICS

Go with the flow

Nature Mater. doi:10.1038/nmat1528 (2005)

Microfluidic systems rely on complex pumps and valves to move tiny quantities of liquids through channels etched into a chip. Luke Lee and his colleagues at the University of California, Berkeley, have developed a way to simplify this set-up.

First, they load the fluid with gold nanoparticles. A low-power laser beam heats the particles at the leading edge of the liquid, which boils and then condenses further up the channel. These fresh droplets coalesce with the bulk of the liquid, effectively dragging it along.

The researchers believe that their approach could be used to make large circuits for manipulating cells and biological molecules.

JOURNAL CLUB

Pier Paolo Pandolfi
Memorial Sloan-Kettering
Cancer Center, New York

We must seek to understand the genetics of cancer susceptibility, argues the director of the Molecular and Developmental Biology Laboratories.

What fascinates me, 30 years after the discovery that cancer has a genetic basis, is that we still know little about the inherited genetic variations that affect our risk of developing the disease.

The discovery of genes that are faulty or mutated within cancer cells has already led to dramatic progress in our ability to treat, and effectively cure, some human cancers.

But only now, with so much more genetic data available, are researchers turning their attention to the inheritable traits that influence our susceptibility to cancers. Importantly, these gene variants could become the targets of preventive medicines.

Recently, Kent Hunter from the National Cancer Institute in Bethesda, Maryland, and his colleagues identified a subtle inherited variation, or polymorphism, in a gene known as *SIPA1* that seems to modify the efficiency with which breast tumours spread (Y.-G. Park *Nature Genet.* **37**, 1055–1062; 2005).

As far as I know, this is the first time an inherited polymorphism has been linked to metastasis — the mechanism by which the tumour spreads and seeds other organs.

Not everyone is convinced that understanding the genetics of cancer susceptibility is a critical goal. But, in 1960, sceptics also questioned the relevance of the finding that patients with a rare form of cancer known as chronic myelogenous leukaemia exhibited a chromosomal abnormality.

It took until 1973, when Janet Rowley and her colleagues proved that the abnormality was caused by the interchange of parts of two chromosomes, for people to realize that this 'Philadelphia chromosome' was the first evidence for the genetic basis of human cancer.

SPECIAL REPORT

Korean scandal will have global fallout

The possibility that Woo Suk Hwang's cloning experiments were faked threatens to undermine confidence in stem-cell research.

In one of the biggest scientific scandals of recent times, South Korea's star cloner Woo Suk Hwang last week asked to retract his landmark paper on the creation of embryonic stem cells from adult human tissue. The request, along with new doubts about his earlier work, confirms what researchers in the field were already starting to realize — that the advance marked by Hwang's research, with all it promised for therapeutic cloning, may amount to nothing.

Worse, scientists fear that the episode will damage not only public perceptions of stem-cell research, but science's image as a whole.

The request for retraction of the paper (W. S. Hwang *et al. Science* **308**, 1777–1783; 2005) came after three authors claimed the work was untrustworthy. Fertility expert Sung Il Roh of MizMedi Hospital in Seoul, claimed on 15 December that Hwang had admitted to him that data were fabricated, and there were no patient-specific cells. In a documentary aired the same day, Sun Jong Kim, formerly of Seoul National University (SNU), told the Seoul-based Munhwa Broadcasting Corporation (MBC) that Hwang had asked him to falsify images. And Gerald Schatten at the University of Pittsburgh asked for his name to be removed from the paper, claiming that information from a team member had caused him to doubt the work's accuracy.

And there are now concerns about earlier work. For example, in the paper in which

Hwang claimed to have extracted the first stem-cell line from a cloned human embryo (W. S. Hwang *et al. Science* **303**, 1669–1674; 2004), figures supposedly showing cloned cell lines are identical to those in an earlier paper showing normal embryonic stem cells (J. H. Park *et al. Molecules and Cells* **17**, 309–315; 2004). *Nature* has also announced an investigation into Hwang's paper on the first cloned dog (see 'Dogged by doubts', page 1059).

Hwang admitted on 16 December that there were errors in the 2005 stem-cell paper, but denied fraud. He maintains that 11 patient-specific stem-cell lines were created as reported, but six were never frozen, and subsequently became contaminated. He says five lines being thawed now will prove his success.

Culture of secrecy

Hwang's claims are meeting with increasing scepticism. "He was given a chance [to explain] but he didn't use it," says a molecular biologist at SNU, who asked not to be named. Robert Lanza of Advanced Cell Technology in Worcester, Massachusetts, who is also attempting to clone human cells, says it is difficult to believe that cell lines of such value weren't stored properly: "What stem-cell scientist doesn't freeze their cells?"

The SNU is investigating the team's work. The lab's atmosphere of pervasive secrecy and tradition of deference towards Hwang will make investigators' job difficult. But if there



At bay: Woo Suk Hwang maintains that further tests will prove his stem cells are genuine.

was fabrication, it will be hard for Hwang to plead ignorance. When *Nature* visited in 2004, he declined to show his first cloned stem-cell line, kept under lock and key. "Many lab members aren't allowed to see it either," he said.

Taken together, the concerns about Hwang's work leave biologists with no proof that stem cells can be extracted from cloned human embryos (see page 1058).

And the scandal's implications will reach further. There have been cases in which fraud has been established that have involved more papers: a 2002 investigation by Bell Laboratories in Murray Hill, New Jersey, found that Jan Hendrik Schön fabricated data in at least 16 papers while working there. But Schön's field of materials science has a lower public

12 February 2004

Woo Suk Hwang of Seoul National University (SNU) in South Korea and colleagues announce in *Science* they have cloned 30 human blastocysts and harvested stem cells from one of them.

6 May 2004

Nature raises ethical questions about Hwang's work after investigations suggest some of the eggs may have come from junior members of the lab. Hwang denies wrongdoing.

19 May 2005

Hwang's team reports in *Science* that it has established 11 embryonic stem-cell lines derived from the skin cells of individuals.

3 August 2005

Hwang and colleagues announce in *Nature* that they have cloned the first dog, an Afghan hound called Snuppy.

19 October 2005

South Korea's government launches the World Stem Cell Hub, an international network for exchanging embryonic stem-cell lines and cloning technology. Hwang is to run it.

12 November 2005

Gerald Schatten of the University of Pittsburgh, co-author on the 2005 stem-cell paper, cuts all ties to Hwang, saying he believes Hwang misrepresented consent issues relating to the 2004 paper. The authors of the 2005 paper correct their data, but say it doesn't alter their conclusions.

TIMELINE OF EVENTS ►



K.-H. KIM/REUTERS

profile than cloning and stem-cell research.

"This is such an important experiment and there was so much publicity around it," says Rudolf Jaenisch, a mouse-cloning expert at the Massachusetts Institute of Technology. "It is shocking to think that it might have been fabricated."

"It will probably affect the general perception of scientists and what we do," says Theodore Friedmann, a gene-therapy researcher at the University of California, San Diego, who has chaired the US Recombinant DNA Advisory Committee. "There's a climate of mistrust of science now that's stronger than in the past. That will be exacerbated by this sort of event."

The debacle may well strengthen the hand of those trying to ban stem-cell research in the United States. "This is an example of the corruption of science that this whole cloning field

has been tending toward, with its end-justifies-the-means mentality," says Gene Tarne of Do No Harm, a Washington DC-based coalition that coordinates opposition to stem-cell research. "For almost a decade now, we've heard these overhyped claims about therapeutic cloning. Somebody took the first step in providing any evidence for these claims and it turns out the evidence simply wasn't there."

Lessons to learn

Researchers are left wondering how such a fiasco happened. The journal *Science*, which published two of Hwang's high-profile papers, has defended its peer-review process. Donald Kennedy, *Science*'s editor-in-chief, says the journal typically takes 120 days to review and publish biology manuscripts. Hwang's 2005 paper took 58 days, leading some to wonder whether it was rushed. "If it's a really hot paper

and you want to get it out quickly, how many shortcuts do you take?" says Nobel laureate Paul Berg of Stanford University, California.

In a press conference on 16 December, Kennedy insisted the journal does not rush papers. "I think we were appropriately suspicious in this case. I don't think this points to a generic fault in the peer-review system," he said.

Asked whether *Nature* could have been caught out in the same way, editor-in-chief Philip Campbell agrees. "We would hope the errors would have been noticed," he says. "But usually reviewers have to take on faith that the authors are presenting what they say they are." He suggests that in future some important claims should be independently tested.

Others are questioning Schatten's role. He promoted the South Korean group in the West, and was senior author on the 2005 paper, although he did not perform any of the experiments it describes. "The lesson I've learned is that I would not be a co-author on a paper unless I was essentially willing to stake my entire career on every piece of data in that paper," says cloning researcher Kevin Eggan of Harvard University in Cambridge, Massachusetts.

Schatten referred *Nature*'s inquiries to Jane Duffield at the University of Pittsburgh Medical Center's news bureau. "He is still not doing interviews with reporters," Duffield wrote in an e-mail.

But some have sympathy for Schatten. "Many scientists would be tempted to do similar things if someone offered them authorship on what seemed like an important breakthrough," says Friedmann.

The field as a whole should tone down its rhetoric, he adds. "I have been very concerned about some of the language used. It seems reminiscent of the gene-therapy experience, where so much promise was obvious, but it was hyped and exaggerated to the detriment of the field. We should be more circumspect." ■

Erika Check and David Cyranoski

Read more on the Hwang case at:

► www.nature.com/news

21 November 2005

Sun Il Roh, a fertility expert at MizMedi Hospital in Seoul and co-author on the 2005 study, admits that 20 of the eggs he gave to Hwang for his 2004 study were paid for. The next day, Munhwa Broadcasting Company (MBC) airs evidence that Hwang used eggs from junior lab members.

1 December 2005

Two review boards say that donations by researchers did not violate ethical guidelines, as they were not coerced. MBC releases results of tests on five samples of Hwang's cell lines. DNA can only be extracted from one line, and MBC says this doesn't match the corresponding tissue sample. Hwang stands by his results.

4 December 2005

Hwang alerts *Science* to wrongly duplicated images published in supplementary information for the 2005 paper. Over the next week, the University of Pittsburgh and SNU announce investigations.

13 December 2005

Science publishes calls for Hwang to cooperate with independent tests. Schatten asks Hwang to retract their 2005 stem-cell paper, claiming allegations from a team member have made him doubt its accuracy. Two days later, news stations across Korea report allegations from Roh that the 2005 paper was faked.

16 December 2005

Hwang says there were mistakes in photographing and preserving his cloned stem-cell lines. He says he will seek to retract the paper, and that he is thawing frozen stem-cell lines to prove the study's validity. *Science* announces that Hwang and Schatten have requested a retraction, but says the entire team's consent is needed.

19 December 2005

Claims circulate that Hwang's 2004 study includes images taken from other papers. *Science* says it is investigating the 2004 and 2005 stem-cell papers. Doubts are also raised about whether Snuppy is really a clone. *Nature* announces an investigation of the Snuppy paper.

Where now for stem-cell cloners?

Scientists are surveying the wreckage left by the debacle involving stem-cell researcher Woo Suk Hwang after three co-authors on his landmark paper said that it could not be trusted. Researchers now face a long slog to rebuild the foundations of their field.

As well as issues relating to trust and public confidence in such a controversial area (see page 1056), the complete loss of confidence in Hwang's work has set the field back by years. It has also taken away what seemed to be firm confirmation of the feasibility of using cloning to produce patient-matched stem cells.

"We thought a fundamental question had been answered," says Alison Murdoch of the University of Newcastle Upon Tyne, UK. "Hwang's results shifted the research focus on to emulating his work. Now we may need to look again at that fundamental step."

Hwang's group claimed two major papers in the past two years that revolutionized the field. In 2004, the group reported that it had cloned a cell obtained from an adult woman (W. S. Hwang *et al. Science* **303**, 1669–1674; 2004). The group claimed it had put DNA from the woman's cell into one of her own eggs, from which the genetic material had been removed. After several days, the egg had developed into a distinct type of early embryo called a blastocyst. From this, Hwang's group supposedly extracted a batch of embryonic stem cells,

potentially capable of developing into any of the body's tissues.

Earlier this year, the group reported that it had vastly improved on this study (W. S. Hwang *et al. Science* **308**, 1777–1783; 2005). The researchers used the same procedure but this time claimed to have transferred genetic material from patients into eggs from unrelated, healthy women, to create blas-

tocysts and extract stem cells. The increased efficiency they claimed also meant that far fewer eggs were needed to create stem-cell lines.

The paper was hailed as a milestone. It apparently provided the first proof of stem cells matched to individual patients and suggested that they were not that difficult to make, confirming the promise of the technique — dubbed "therapeutic cloning" — for producing replacement cells and tissues. It also seemed to settle lingering questions about whether cloning actually worked. Many scientists had not been convinced by the results of Hwang's 2004 experiments. Because the egg and donor DNA came from the same person, it was impossible to be sure that the stem-cell line was created from the donor cell instead of the egg.

In the past few days, doubts have also been raised about the authenticity of the 2004 paper (see page 1056). But whether it is valid or not,

the loss of confidence in the 2005 study leaves scientists with no proof that adult cells can be cloned — let alone used to produce stem cells. "Hwang's work gave people confidence to move into this difficult area," says Alan Colman, head of Singapore-based regenerative-medicine company ES Cell International and a member of the team that cloned Dolly. "But maybe it's harder than we thought."

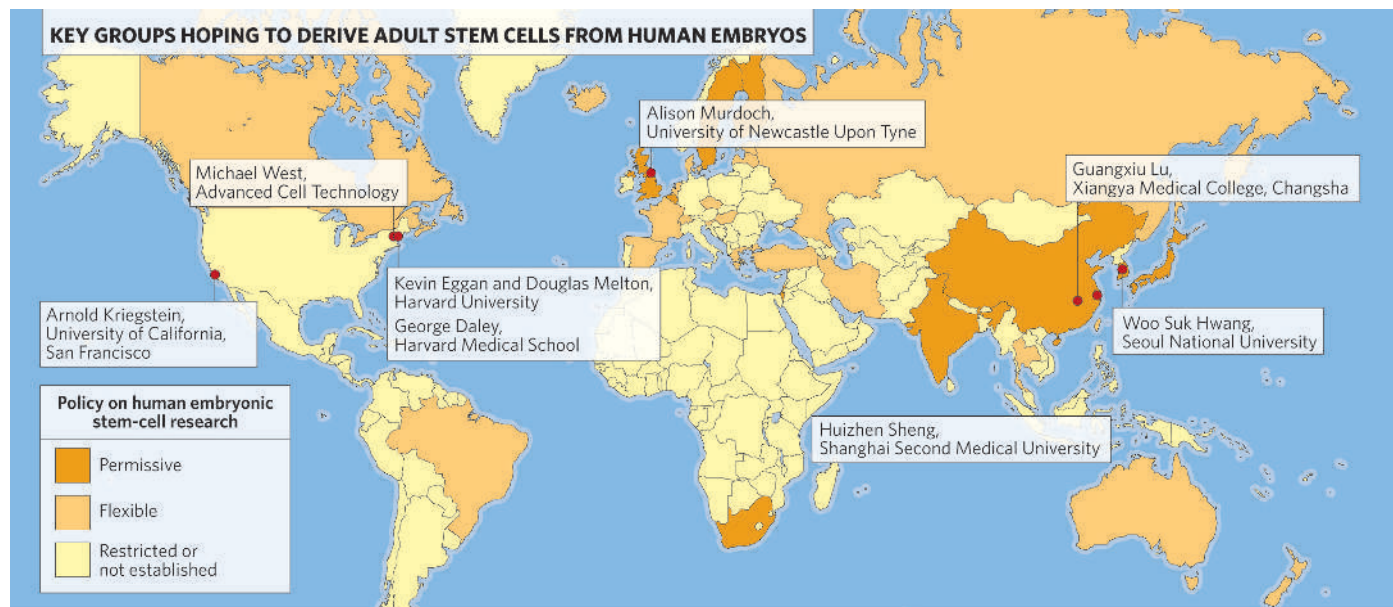
"We're back to knowing that animal cloning is possible but wondering whether it is possible in humans," adds Kevin Eggan of Harvard University in Cambridge, Massachusetts. "This is an enormous setback."

With Hwang's work set aside, results from other groups are sparse. After Hwang's apparent success, researchers flocked to his lab to learn his methods, but most, such as Eggan, and George Daley of Harvard Medical School, are still waiting to get approval to use them in their home countries.

There have been a few baby steps, however. In 2001, a company called Advanced Cell Technology (ACT), based in Worcester, Massachusetts, described its attempt to create cloned human blastocysts. But the group's clones survived only a few days and never made it to the blastocyst stage (J. B. Cibelli *et al. J. Regen. Med.* **2**, 25–31; 2001). The researchers abandoned their work because of lack of funding once Hwang claimed success.

In 2002, Chinese researchers made headlines with a report that Guangxiu Lu of the Xiangya

"We're back to wondering whether cloning can be done in humans. This is an enormous setback."



Dogged by doubts

Is Snuppy really a clone? With the credibility of his creator Woo Suk Hwang under fire, the dog's credentials are being challenged.

The Afghan hound was supposedly the first dog to be cloned (B. C. Lee *et al. Nature* **436**, 641; 2005). Cloning dogs presents unusual challenges because, compared with other mammals, the egg cells are difficult to mature *in vitro*. Hwang's group says it used the same technology as in its human experiments — removing the nucleus from a donor's cell and inserting it into an egg cell, a process called somatic-cell nuclear transfer (SCNT).

But Robert Lanza, a stem-cell expert at Advanced Cell Technology in Worcester, Massachusetts, and a competitor with Hwang in human therapeutic cloning, says the paper should now be seriously re-examined.

Lanza says that Snuppy, seen on the right with the dog from which he was supposedly cloned, might have been created by a technique called embryo splitting, in which cells from an early-stage embryo are divided and then implanted separately. The technique creates identical twins. One set of cells could have been used immediately to create a dog while another was frozen and stored. If the frozen cells were later used to create a dog with identical DNA, that could be presented as an SCNT clone.



SEOUL NAT'L UNIV./GETTY IMAGES

Such trickery could be caught by examining mitochondrial DNA, which is passed maternally with the egg cell. If Snuppy were really a SCNT clone, he should have the mitochondrial DNA of the dog from which the egg was taken. If he's a fake, he'd share it with the dog from which he was supposedly cloned.

Mitochondrial DNA data have not been part of previous cloning papers, and were not presented in *Nature*. Lanza suggests that it would now be a good idea to do the test. "If the mitochondrial DNA is the same, that's the end of that paper," says Lanza.

Nature is starting an investigation, including a mitochondrial DNA test, that is unlikely to be complete before January 2006.

David Cyranoski

Medical College in Changsha, Hunan, had cloned human blastocysts from adult cells (*Chinese Sci. Bull.* **48**, 1840–1843; 2003), although she had not been able to extract stem cells from any of them. Also, Huizhen Sheng of Shanghai Second Medical University claimed to have extracted stem cells from embryos created by introducing adult human DNA into rabbit eggs stripped of their own chromosomes (Y. Chen *et al. Cell Res.* **13**, 251–263; 2003).

And in August, Murdoch's group reported the creation of a single blastocyst from a cloned cell (M. Stojkovic *et al. Reprod. BioMed. Online* **11**, 226–231; 2005). The blastocyst died before yielding any stem cells. And as the cloned cell was itself an embryonic stem cell, the paper does not show a way of making stem cells matched to adult patients from scratch.

Murdoch says she does not relish now being a leader in the field. "I'm not interested in striving to be the first to get somewhere," she says. "The problems in South Korea highlight the difficulties in racing to get results."

She also laments the rules and regulations that many scientists think have hamstrung stem-cell research (see map, opposite). "The more people who are working on this the better," she says. "But the fundamental problem

is that it is banned in so many countries."

But researchers in the field are hopeful that progress can be made. "This needs to be done right," says Michael West of ACT. "And many of us are determined to make it happen." He says his company now plans to revisit the work. Eggan and Douglas Melton, also at Harvard University, hope to get approval from the review boards that oversee their research in time to start work cloning human embryos early next year. Daley is planning experiments similar to those done by Murdoch's group. And Arnold Kriegstein and his group at the University of California, San Francisco, plan to try to replicate Hwang's methods with their own materials.

But for others, the episode merely confirms that therapeutic cloning is not the way forward. "I always had my doubts about therapeutic cloning to generate patient-matched cells," says Stephen Minger, a stem-cell researcher at the Wolfson Centre for Age Related Diseases in London, UK. He believes that banking stem-cell lines from normal embryos, so that they can be matched to patients once they are made, is a more realistic prospect.

Erika Check

Additional reporting by Tom Simonite and Carina Dennis

India makes waves over tsunami warning system

HYDERABAD

India has agreed to share seismic data from four of its monitoring stations as part of a tsunami warning system for the Indian Ocean. But its offer has left many unimpressed.

The warning system will use a maze of deep ocean sensors and tide gauges surrounding the fault that ruptured on 26 December 2004. This earthquake triggered a tsunami that killed more than 200,000 people in 11 countries. But crucial to the network will be real-time seismic data from stations in the region.

India has been averse to sharing its seismic data in order to keep information about its underground nuclear tests a secret. "The only station that is available to the global seismic network has a delay of about three weeks before data are disseminated," says Walter Mooney of the US Geological Survey, headquartered in Reston, Virginia.

India's offer, announced at the second meeting of the Intergovernmental Coordination Group (ICG) in Hyderabad last week, is limited to data on earthquakes with a

magnitude of six and above, along the coast of Indonesia and Pakistan. Signals from nuclear tests would be much weaker than this. "For the purpose of tsunami warning we think our offer should be quite satisfactory," India's science secretary Valangiman Ramamurthy told *Nature*.

Not everyone agrees, because of the time it would presumably take to filter the data. "A delay of even a minute in the dissemination of earthquake information could increase casualties," says a report by an ICG working group released at the meeting. "We were pinning our hopes on real-time seismic signals from India," adds Reinhold Ollig, head of a delegation from Germany that is helping Indonesia to build a national tsunami-warning centre in Jakarta. "Now we may have to upgrade a station in Sri Lanka for a real-time link."

The Indian offer, even though it is limited, is "a sign of progress," Patricio Bernal, executive secretary of the United Nations' Intergovernmental Oceanographic Commission, told *Nature*. He says Indian Ocean countries

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A tsunami early-warning network for the Indian Ocean should be in place by 2009.

should be in a position to confirm the advance existence of a tsunami by September 2006, and that the fully fledged warning system is on track for completion by the end of 2008.

But as the first anniversary of the Asian

REUTERS/S. CRISP

Diet book attacked for its high-protein advice

SYDNEY

A diet book developed by researchers at Australia's largest government laboratory network has already made the organization more than A\$1.5 million (US\$1.1 million) in royalties. But its success is feeding a growing body of critics who say that its high-protein message is not supported by the evidence. They also question the influence of the meat industry, which sponsored it.

"It's far more successful than we ever anticipated," says Manny Noakes of the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Adelaide, who wrote *The CSIRO Total Wellbeing Diet* with her colleague Peter Clifton.

The book has become a national bestseller, having sold nearly half a million copies since its launch in May. It went on sale in Britain in September, with release in further countries, including the United States, planned for 2006. Its recommendations even feature on



Manny Noakes (left) co-authored a book that advocates eating more protein for weight loss.

the menu at Australia's Parliament House.

But critics have spoken out about the possible influence of the Australian meat and livestock industry, which funded a large portion of the research behind the high-

protein diet. "There is a bias towards the sponsor's product which is not justified by the results of their research," says Rosemary Stanton, a nutritionist and visiting fellow at the University of New South Wales in Sydney.

The diet advocates a much higher protein intake than that recommended by most national guidelines. People on a typical Western diet obtain about 15% of their energy intake as protein, but the CSIRO diet recommends doubling that to 30–35% while reducing carbohydrate intake. To achieve this, Noakes and Clifton suggest eating more meat and fish at lunch and dinner.

"You have to ask why they didn't promote more plant-based proteins," says Stanton. "Did their choice of protein come about because of the sponsor?" The authors insist that the industrial sponsors were kept at arm's length. "They didn't have any impact on the design of the study and how we interpreted the results," says Clifton.

AAP/A. PORRITT

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

tsunami approaches, there is concern at the news that the ICG has dropped the idea of one or two countries being responsible for issuing a warning across the region through the network. The ICG was worried that the proposal had "overly controlling connotations", despite a similar system being in use at the Pacific Tsunami Warning Center in Hawaii. Instead,

Nonetheless, Meat and Livestock Australia, which represents the nation's livestock industry, has been a keen publicist: it distributed a booklet on the diet in a women's magazine. This was noticed by the publisher Penguin, which then commissioned the book. Royalties go to CSIRO nutrition research.

Other scientists are concerned that the evidence behind the diet is weak, and that by putting its name to the book the CSIRO is giving the diet unwarranted credibility. "The CSIRO name unquestionably sells more copies," says Jim Mann, a nutrition expert at the University of Otago in Dunedin, New Zealand. "But the hype goes beyond what the research proves."

"The main trial showed no difference in weight loss compared with a conventional diet," points out Patrick Holford, founder of the Institute for Optimum Nutrition based near London, UK. He believes that sticking to such a diet could elevate the risks of breast and prostate cancer, stress the kidneys and adversely affect bone mass.

the ICG suggests that the Intergovernmental Oceanographic Commission should accredit certain nations as 'watch providers' from whom, under bilateral agreements, other nations could obtain details of any events detected. It would then be up to individual nations to decide whether to issue a warning within their own territory. Indonesia, India, Thailand, Malaysia and Australia plan to have their national warning centres in operation before 2009.

Individual nations will be able to enter into bilateral arrangements with as many watch providers as they wish, which means that there will not be a single alert but several voices, depending on how many providers each nation ties up with. "There is going to be chaos," warns K. Radhakrishnan, former director of the Indian National Centre for Ocean Information Services in Hyderabad.

Here, too, India is choosing its own path. It is investing US\$30 million to upgrade its 70 seismic stations, deploying ten deep underwater pressure sensors and installing 50 satellite-linked tide gauges. It plans to have its warning centre running by September 2007 but says it will not subject itself to the ICG's accreditation process. "What India is doing is adequate for the entire Indian Ocean region," says Ramamurthy. "If any country wants to work with us in tandem we have no problem."

K. S. Jayaraman

"I think it is dangerous long-term," he says.

The authors based the diet on several studies, the largest being their own trial of 100 overweight women over 12 weeks (M. Noakes *et al. Am. J. Clin. Nutr.* 81, 1298–1306; 2005). Half the women were given a high-protein diet and the other half a high-carbohydrate diet. Both diets contained the same number of calories, and both groups of women lost the same amount of weight. But the authors say their recommendations are valid because women with high triglyceride levels — a marker of insulin resistance — shed significantly more weight on the high-protein diet. Participants were also more likely to stick with the high-protein diet.

The CSIRO stands by its decision to commercialize the research. "The CSIRO has always published books on its scientific work and put its name to publications, and this is no exception," says a spokeswoman. "The decision to publish was in response to many consumers asking for further details of the diet."

Carina Dennis

ON THE RECORD

"I hate being cold."

British swimmer Lewis Pugh has some qualms about attempting a record 1-kilometre swim in the frigid waters of the Southern Ocean.

"On a deeper level Barbie has become inanimate... This may go some way towards explaining the violence and torture."

Psychologist Agnes Nairn explains how young girls apparently see the ubiquitous plastic doll as a symbol of excess, triggering them to decapitate and maim their Barbies.

Sources: Reuters, The Times

SCORECARD



Pygmy elephants

Miniature pachyderms in Borneo get some big attention, in the form of global positioning system collars that track their every move through the rainforest.



Sounds without words

An obscure buzzing sound present in some 70 African languages, and known as the labiodental flap, joins the International Phonetic Alphabet — the first such addition in a dozen years.



Army hygiene

Studies of body lice and dental pulp from French soldiers buried in Russia suggest that many in Napoleon's army suffered from louse-borne diseases, including typhus and trench fever.

NUMBER CRUNCH

\$54 is the cost of a 'custom star kit' through one of the many star-registry agencies that advertise buying a star in the name of a loved one — a perfect Christmas gift.

1 million people have signed up.

0 is the number of privately purchased star names recognized by the International Astronomical Union, the organization in charge of naming celestial objects.

The heat was on in 2005

As 2005 draws to a close, climate scientists are making their annual pronouncements on how its temperatures compare to historical records. And although this year is among the warmest ever recorded, small differences in the claims highlight the uncertainty of such rankings.

Depending on whom one believes, 2005 will end up just above or below 1998 as the hottest year on record. Most significant, climate scientists say, is that this year's readings occurred without the help of a major El Niño event. "In just seven years, the background global temperature has increased to a level equal to the peak in the 1997–98 El Niño," says James Hansen, a researcher at NASA's Goddard Institute for Space Studies in New York City.

That record-breaking El Niño slathered the tropical Pacific with anomalously warm sea water. There was no such event this year, but many other regions were notably warm — including the North Atlantic, where an unprecedented number of tropical cyclones formed.

Hansen says that NASA is likely to dub 2005 as the warmest year on record, but a team at the University of East Anglia in Norwich, UK, is poised to rate it as number two, behind 1998. And a preliminary report from the National Oceanic and Atmospheric Administration (NOAA) shows a photo finish between the two years, with 1998 ahead by a nose (see 'Sources of disagreement'). Final rankings will be released over the next few weeks.

This year's heat was not a total surprise — NASA predicted early in 2005 that it would be one of the warmest years on record. Over the past century, says NASA, Earth's average surface temperature has risen 0.8 °C, with three-quarters of that occurring since the 1970s.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

This year's record-breaking temperatures included a devastating heatwave in Pakistan.

Nine of the ten warmest years on record have occurred since 1995.

Hansen, who compiles the annual rankings for NASA, says the recent warming is consistent with the increase in heat-trapping greenhouse gases in the atmosphere. "Climate change is real and should begin to be noticed by real people," he says.

Although differing rankings for 2005 might puzzle the public, it is less of an issue for the scientists who compile them. Most of the time, the ratings agree. "People sometimes make too

much of whether a year is ranked warmest or second warmest," says Jay Lawrimore, who oversees month-to-month tracking for NOAA.

Scientists hope to put the rankings in better perspective by pointing out uncertainties in them. In 2006, NOAA will shift to an analysis technique that will include uncertainty ranges for the first time. This may reduce the drama of the year-end rankings, but it could also accentuate just how many of the past few years lie at the top of the temperature heap. ■

Robert Henson

Sources of disagreement

There are three teams that rank global temperatures. Their results vary mainly because of differences in how they combine data sets.

Each group draws on a different mix of the planet's land-based temperature stations to construct a temperature record. The University of East Anglia's Climatic Research Unit (CRU) uses about 4,200 stations worldwide; the National Oceanic and Atmospheric Administration (NOAA) uses 7,200 and NASA uses 6,000.

They also differ in how they

analyse this information. NASA and NOAA pool their data, weighted by area, across the globe. But the Northern Hemisphere has much more land than the Southern: "We think this adds a northern bias," says Philip Jones of the CRU. His team averages the data for each hemisphere, then combines them. Another difference is that NASA calculates its temperature differences using a 1951–80 base period; the others use 1961–90.

But overall, the results are more

alike than they are different. The three groups report similar rates of warming over land in the past century, according to a recent analysis by NOAA's Russell Vose.

Adding measurements from the ocean brings more uncertainty. For decades, scientists relied on fairly crude sea-surface-temperature measurements collected by ships through buckets and engine intakes. But by the early 1990s, sea-surface data from ships and buoys became more widely available, as did air temperatures

construed from satellite data.

NOAA and NASA use an index that includes all these ocean sources; the CRU and the Hadley Centre for Climate Prediction and Research in Exeter, UK, rely on ship and buoy data. There is no consistent difference in the results, says Hadley's John Kennedy, but this year the CRU/Hadley index pegs ocean temperatures as being cooler than they were in 1998. That may be why that team seems likely to place global air temperatures short of the 1998 record. **R.H.**

Australian report calls for relaxation of stem-cell laws

An Australian report on embryonic stem-cell law could pave the way for the creation of one of the world's most permissive research environments.

The government-commissioned report, which was released on 19 December, recommends relaxing the country's current ban on therapeutic cloning and establishing a national stem-cell bank. It also advises permitting the creation of cross-species chimaeras — where human somatic cells are fused with non-human eggs — for research and training.

Australia's current laws, dating from 2002, were seen as an impediment to the field; some of the nation's leading researchers and biotechnology entrepreneurs have already moved elsewhere. "If the government adopts the recommendations, we would be in a very attractive position to get back some scientists," says Alan Trounson, a reproductive biologist at Monash University in Melbourne.

DuPont fined over safety data on Teflon chemical

Chemical titan DuPont will pay the largest-ever civil administrative penalty levied by the US Environmental Protection Agency. The fine — \$10.25 million — relates to alleged infractions concerning the chemical perfluorooctanoic acid (PFOA).

PFOA, a possible carcinogen, is used to make several DuPont products, including Teflon non-stick coating for cookware. Most of the charges involve the company's failure to inform the agency about data on the chemical's risks.

Under the 14 December agreement, DuPont will also spend \$6.25 million on related research, such as determining whether and how any of its products can break down to form PFOA. Further funds will go towards a green-chemistry project in schools in Wood County, West Virginia, the site of one of DuPont's chemical plants.

Innovation act proposes big boost for US research

Legislation introduced in the US Congress last week proposes a doubling of the National Science Foundation's budget between 2007 and 2011.

Among its provisions, the bill would set up a grants programme to fund innovative but high-risk projects, allocate nearly \$100 million a year for graduate research grants, and encourage the development of

US coastguard urged to take back ageing icebreakers

The US Coast Guard, not the National Science Foundation (NSF), should run the country's ageing fleet of polar icebreakers, according to an interim report released on 14 December by the National Research Council.



The NSF took control of the three US icebreakers used for research from the coastguard earlier this year. But the ships will need expensive repairs, and there is little experience at the NSF in running them.

"We as a group agreed that it really didn't make sense to have the NSF in charge of these icebreakers," says committee member Julie Brigham-Grette of the University of Massachusetts, Amherst.

Two of the ships, which are 30 years old, usually help break a channel to McMurdo Station in Antarctica, but this year the NSF also chartered a newer Russian icebreaker. The report urges that at least one US vessel should be capable of clearing the way to McMurdo each year.

Climate change is expected to make the Arctic Ocean more accessible to shipping. Icebreakers might be needed there in an emergency, or to break the way for commercial traffic, increasing the need for a reliable fleet, the report says.

US COAST GUARD

regional hot spots for technological innovation.

The act is a response to a report released a year ago by the Council on Competitiveness, a group of US chief executives, university presidents and labour leaders. The report called for investment in education, training, research and development, and commercialization of research to keep the United States globally competitive.

Even the bill's supporters do not expect all of it to pass, but hope that pieces of the proposal may become law.

DaimlerChrysler tops league of R&D spenders

The German car maker DaimlerChrysler is the world's biggest private investor in research and development, with an annual spend of €5.6 billion (US\$6.7 billion), according to a European Commission study published on 9 December.

Sixteen companies — seven in the United States, three each in Germany and Japan and one each in Britain, France and Finland — have an annual research and development budget bigger than the European Union's €3.5-billion Framework Programme for research. The European Commission has proposed doubling the programme's budget between 2007 and 2013.

The US pharmaceutical company Pfizer comes close behind DaimlerChrysler, followed by the US Ford Motor Company, Japan's Toyota Motor Company, and Germany's Siemens. The British company GlaxoSmithKline ranks at number 11 with €4.01 billion.

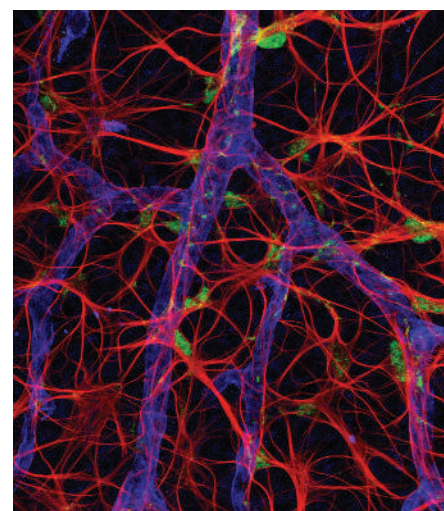
Image of retina takes prize for visions of biology

Blue blood vessels combine with red astrocytes in this award-winning image of an ageing rat's retina.

In a competition sponsored by optics firm Olympus, judges cited the image's combination of technical accomplishment, beauty and scientific significance in awarding top prize to Hussein Mansour, a graduate student at the University of Sydney, Australia. Studies of astrocytes in ageing animals could shed light on how the human brain deteriorates as it gets older.

Other winners included images of fly testis cells and the wings of a moth.

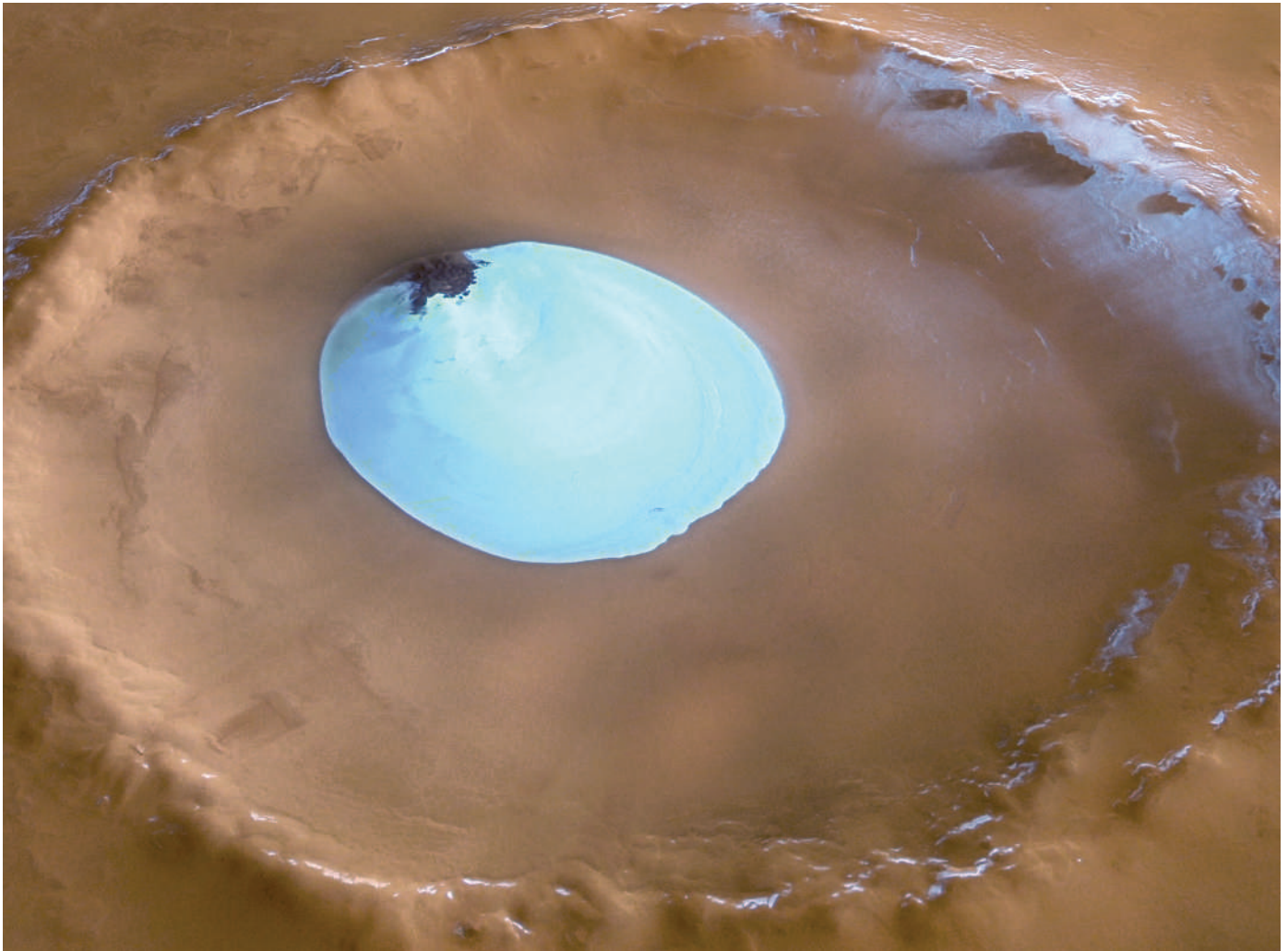
► www.olympusbioscapes.com/gallery/2005/index.html



Astrocytes are red, blood vessels are blue: the rat brain could shed light on human ageing.

H. MANSOUR

2005 GALLERY



G. NEUKUM/ESA/DLR/FU BERLIN

FIRST GLIMPSE...

It can sometimes seem as if all the great discoveries in science must have been made already. And yet, every year, we get our first glimpses of things — creatures, heavenly bodies, states of matter or molecules — that give pause for thought. In this last issue of the year, *Nature* presents a gallery of such wonders — a few of our favourite images from 2005.

Many of these pictures accompanied scientific papers, but they have a power that academic prose cannot touch. We humans often don't believe something until we see it. Here, then, are ten more things to believe in. Some are rough shots, taken on the run; others are more like considered artwork, such as the breathtaking images from space.

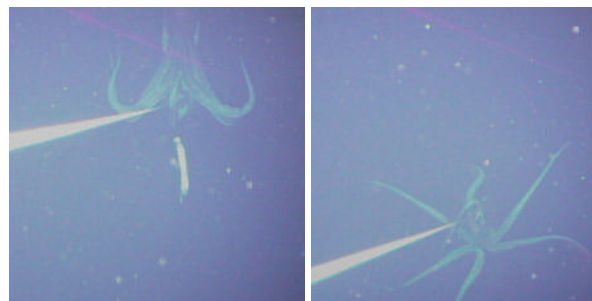
Researched and written by Emma Marris.

MARS

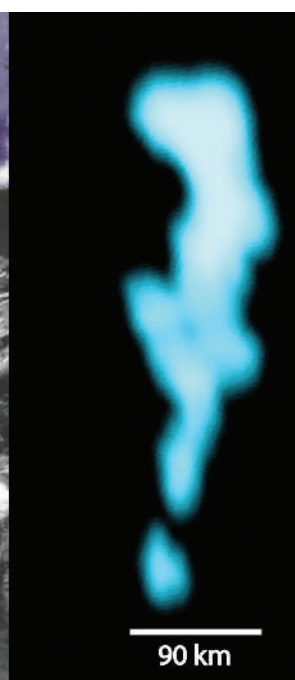
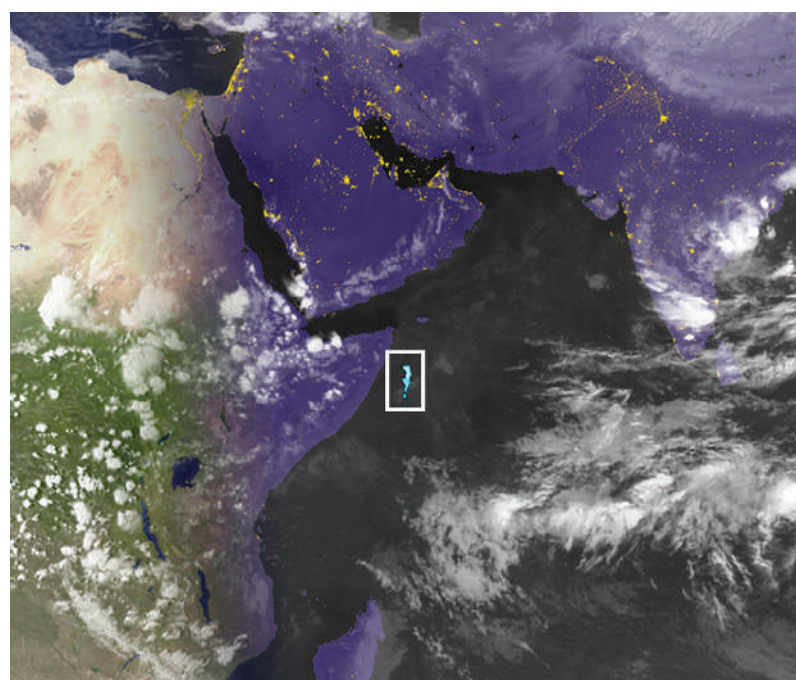
What gives this image of a frozen lake on Mars, taken by the European Space Agency's Mars Express orbiter, its air of mystery? The colours, which have been added to the original black-and-white image? Or the spray of white along the lip of the crater, which looks like early morning frost?

MONSTER FROM THE DEEP

At last, the giant squid (*Architeuthis*) is photographed alive. Japanese researchers lured this eight-metre specimen with a baited line. Tsunemi Kubodera of the National Science Museum, and Kyoichi Mori of the Ogasawara Whale Watching Association, both in Tokyo, took the shot, and a 5.5-metre-long piece of the creature's arm, which became tangled in the line. "We were so excited that we could not stop shouting 'We have hooked a giant squid!'" says Kubodera.



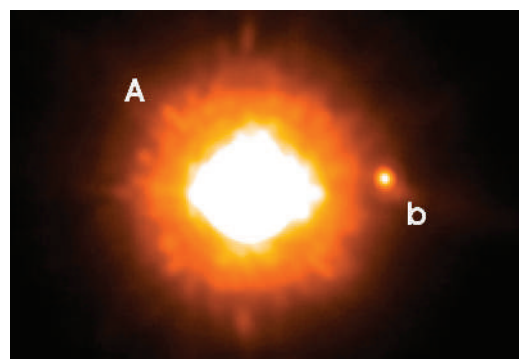
T. KUBODERA & K. MORI PROC. R. SOC. LOND. B 272, 2583–2586 (2005)



MILKY SEA

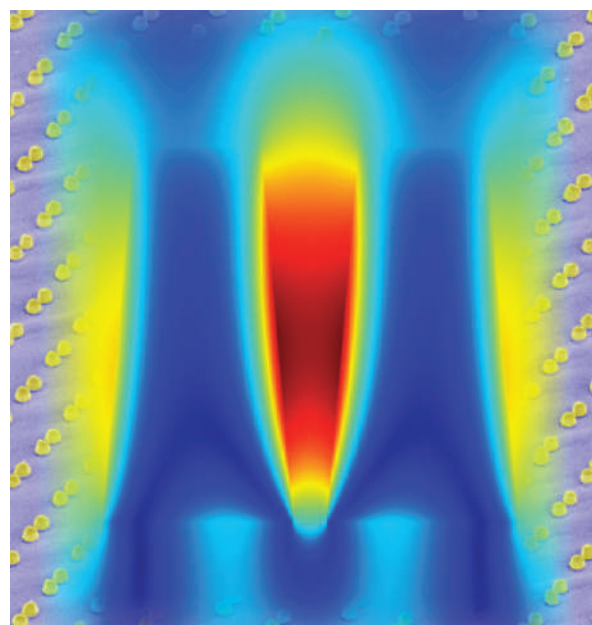
Maritime lore was doubly vindicated this year. Seafloor sensors confirmed that ship-swamping 'rogue waves' really do exist, and this satellite image offered support for the 'milky seas' of legend. The Connecticut-sized glowing smudge, first spotted by a ship in the Indian Ocean, is thought to be made by bioluminescent bacteria. The picture was tracked down by the US Naval Research Laboratory in Monterey, California.

S. D. MILLER ET AL. PROC. NATL. ACAD. SCI. USA 102, 14181–14184 (2005)/NRL



LIGHT SHOW

Alexander Grigorenko's lab at the University of Manchester, UK, got halfway to making a perfect lens, which would reflect no light. The blue bits in this image are areas where the magnetic component of light is not reflected, thanks to an arrangement of tiny gold pillars.



A. GRIGORENKO ET AL. NATURE 438, 335–338 (2005)

PLANET HUNT

The dot on the right of this image (b) could be the first photo of an extrasolar planet. Orbiting a sun 400 light years away called GQ Lupi, the planet is thought to be bigger than Jupiter. It is three times farther from its star than Neptune is from the Sun, giving it an orbital period of 1,200 Earth years. A group led by Ralph Neuhauser, at the Astrophysical Institute and University Observatory in Jena, Germany, captured this image of reflected glory.

R. NEUHAUSER ET AL. ASTRON. ASTROPHYS. 435, L13–L16 (2005)/ESO

A. LEVSKAYA ET AL. NATURE 438, 441–442 (2005)



◀ A PORTRAIT

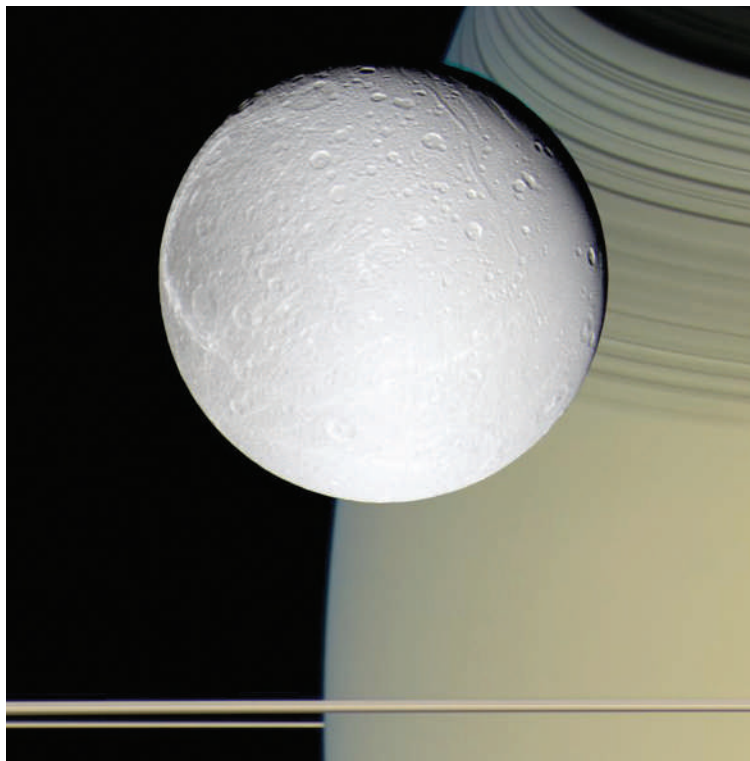
Reminiscent of an early daguerreotype, this image is made of — and by — bacteria. The *Escherichia coli* have been genetically modified both to detect light and to switch off the production of a dark pigment in response. A team consisting mostly of students from the University of Texas, Austin, and the University of California, San Francisco, made this image of Andrew Ellington, one of their professors.



FULL MOON

Since the Cassini-Huygens mission arrived at Saturn last year, NASA's Cassini orbiter has circled the ringed giant, sending back stunning pictures, and the Huygens probe has dropped to the surface of the moon Titan. These Cassini snaps of another moon, Dione, show its icy surface in unprecedented detail, with the shadows of Saturn's rings projected on to the planet behind it.

NASA/JPL/SPACE SCIENCE INST.



LITTLE SWEEP ▶

This brush of carbon nanotubes weighs just 50 micrograms, and can paint the inner surface of a tube 300 micrometres wide — twice the width of a human hair. The first of its kind, the broom was made by Anyun Cao, from the Rensselaer Polytechnic Institute in Troy, New York, and his team.



A. CAO ET AL. NATURE MATER. 4, 540–545 (2005)



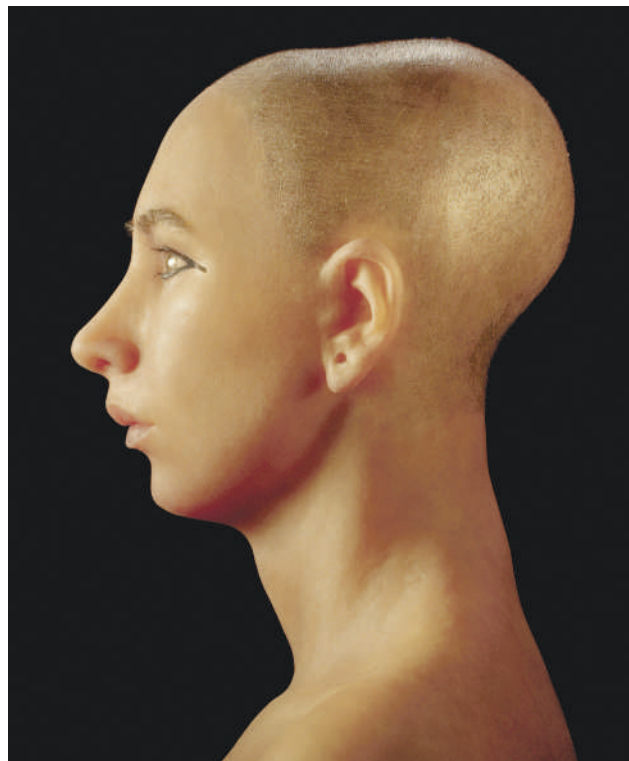
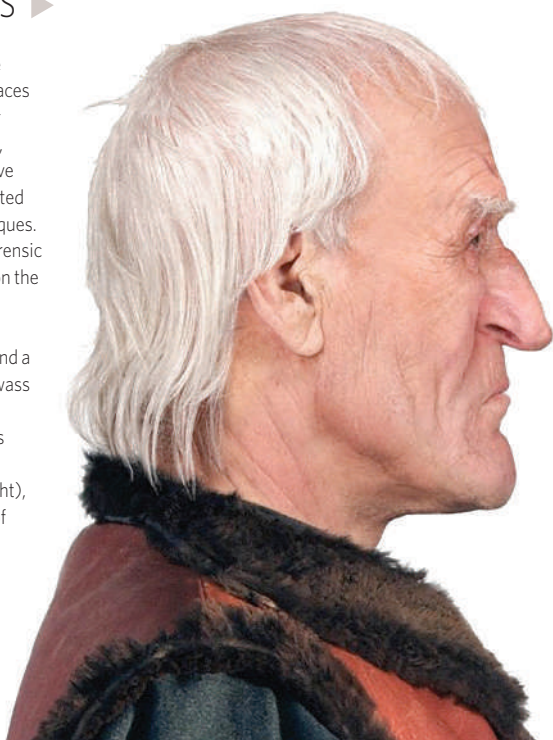
T. BREUER ET AL. / PLOSBIOL. 3, E380 (2005)

STICK WITH IT

Gorillas in captivity are known to use tools. This year, the same was shown for wild apes. Two female gorillas were spotted using a branch as a depth-finder and bridge. Thomas Breuer of the Wildlife Conservation Society in New York and team members caught them on camera.

TWO FACES

One prince among astronomers and one earthly king, whose faces have gone unseen for 462 and 3,300 years, respectively. Both have now been reconstructed using forensic techniques. The Polish police's forensic laboratory put flesh on the bones of Copernicus (right), which were exhumed this year. And a team led by Zahi Hawass of Egypt's Supreme Council of Antiquities did the same for King Tutankhamen (far right), following a CT scan of the mummy.



STR/AFP/GETTY IMAGES; E. DAYNES/NATL GEOGRAPHIC IMAGE COLLECTION

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

THE PARTY GENE

In the first of three
Features looking at
aspects of alcohol, **Siëlle
Gramser** discovers how
yeast first opened the
floodgates of intoxication.

Steven Benner jokingly calls himself a dilettante. A biochemist at the University of Florida in Gainesville, Benner dabbles in a wide range of disciplines, from bioinformatics to astrobiology. His aim is to gain insight into the basic chemical rules that govern how life works — both here and, ultimately, on other planets. But although science drew his gaze to the skies, it was alcohol that brought him back down to Earth. Or, to be more exact, the enzymes that can both make and consume it.

Alcohol dehydrogenase is best known as the enzyme that breaks down alcohol in the body, and as such it has been studied exhaustively. But Benner and other researchers in the field have now turned to its evolution, and their work is providing fresh insight into the puzzle of why some creatures, such as yeast, came to make alcohol and why so many others, including ourselves, can tolerate it.

Alcohol dehydrogenase — ADH for short — is a blanket term applied to a large and diverse group of enzymes. In many creatures, including ourselves, they help to convert alcohols, such as ethanol, into compounds that other enzymes can break down and extract energy from. But in a number of microorganisms, they can help the reverse reaction, making alcohols as part of the process of extracting energy from sugars.

The stars of these alcohol-producers are the yeasts. Not only do *Saccharomyces* species of yeast churn out oodles of ethanol, they can also tolerate far higher concentrations of it than other microorganisms. Brewer's yeast (*S. cerevisiae*) owes this ability to two alcohol dehydrogenases: ADH1, which makes ethanol, and ADH2, which breaks it down for

use as an energy source. Yeast not only brews its own moonshine, it consumes it too — “to the last drop”, as Benner says.

At first sight, this makes no sense. Making ethanol from sugar and then consuming it is energetically far more wasteful than simply consuming the sugar. Researchers have long pondered why yeast goes to all that trouble. Although it might be nice to think that there is a creature out there whose *raison d'être* is to party, evolution doesn't work that way.

Make or break

Benner and his team came across the explanation when hunting for the origins of ADH in yeast. Benner is interested in combining the study of genes and proteins with geology and palaeontology to gain insight into the history of life on Earth and present-day protein function. “Every biomolecule is better understood if we know its history as well as its structure,” he says.

The ADH genes in yeast make an intriguing subject for this approach. When yeast gained its ability to make alcohol, it must have done so

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Still life: why did Brewer's yeast
evolve the bizarre trick of
producing alcohol?

as a result of a selection pressure in its environment and, what is more, this would have had a knock-on effect on other creatures. So working out when and how the ADH enzymes came to be could open a small window onto what ecosystems were like back then.

ADH genes and the proteins they make are well studied and have been isolated from many different species of yeast, so Benner's team had plenty of useful material to work with. The goal was to reconstruct the original gene that was duplicated to give rise to *ADH1* and *ADH2*, and to ask what its function was — did it make alcohol, or did it break it down?

From a database of the sequences of related ADH genes in various yeasts — combined with additional ADH genes specially sequenced for this study — Benner and his colleagues assembled an evolutionary tree of yeast ADH. This showed where the ancestral gene would have fitted in and helped the researchers work out its most likely amino-acid sequence. Inferring the past from the present isn't perfect, so they ended up with 12 slightly different candidate genes¹.

Fruitful collaboration

The group then reconstructed all 12 genes and tested them in yeast to see how the enzymes they produced compared with today's ADH enzymes. The supposed ancestor turned out to be most similar to modern-day *ADH1*, the one that helps yeast make alcohol.

The same evolutionary tree helped the team to estimate when the ancestor gave rise to the two present ADH genes. This information offers some insight into what drove the strategy. Was it humans breeding yeasts and selecting them to accumulate alcohol? Or did the event take place long before that?

The group found that duplication of the ancestral gene took place between 80 million and 60 million years ago, which means that humans could not have had anything to do with it. Rather, Benner thinks it was down to flowering plants. "The hypothesis is that it occurred near the time Earth first provided yeast with fleshy fruits," he says. With their temptingly large amounts of sugar, the fruit called for a clever strategy. "Yeast 'realized' there was a lifestyle opportunity, which involved making large amounts of alcohol as a way of defending the resources against competing organisms," Benner explains.

In other words, yeast came up with a way of 'pickling' the fruit by producing alcohol, which would have made the fruit toxic to its competitors. This had a knock-on effect on its wider ecosystem: as well as killing off its competitors, yeast had created a niche in fermenting fruit for any organism that could devise a way to cope with the alcohol.

It was around this time that the fruitflies emerged. Feeding on yeast and fruit juices in rotting fruit that can easily contain alcohol concentrations of 4% or more (about the same as beer), the fruitfly (*Drosophila*) and its larvae

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Juicy fruits: there has been an evolutionary arms race to hog the sugary treats of flowering plants.

found themselves in need of a mechanism for breaking down alcohol. *Drosophila* came up with its own form of ADH, structurally unrelated to that of mammals and yeast. In fruitflies, ADH plays a role in alcohol tolerance but also in energy metabolism, allowing the fly to use alcohol — indeed many different alcohols — as energy sources.

Different species of *Drosophila* live on different fruits, which in turn produce different combinations of alcohols when they ferment. Given that the biology of ADH is well understood, and that fruitflies are ideal for doing genetics studies, scientists have turned to studying the enzyme to understand how natural selection shapes it to prefer different alcohols in different species. Such studies provide an elegant link between a creature's ecology and the molecular changes that allowed it to exploit its niche.

Luciano Matzkin, an evolutionary biologist at the University of Arizona in Tucson, recently looked at ADH in two species of *Drosophila* that feed on different plants. He compared the different versions of the *Adh* gene in each fly, and identified key changes to the enzymes' structures that could have helped

the flies adapt to different alcohols².

Although alcohol tolerance is clearly an important trait for fruitflies, it is not the only function ADH seems to have in *Drosophila*. "It has played various roles during the evolution of the fruitfly," Matzkin points out. "It pops up in many different places." One of these is related to how well flies can resist a hot environment. Different populations of flies living at different latitudes have different versions of the *Adh* gene. And these patterns can shift rapidly in response to climate change, giving scientists a ringside seat for watching evolution at work, as well as a way of seeing the effects of global warming on ecosystems.

Rapid response

Together with others, Ary Hoffmann, evolutionary geneticist at La Trobe University near Melbourne, Australia, found that a particular version of the *Adh* gene, called *Adh^S*, in Australia has spread south by some 400 kilometres in only 20 years³. This version of the gene is associated with heat resistance. "Twenty years is rapid in evolutionary terms," Hoffmann points out. The speed of change suggests that different versions of *Adh* can make a big difference to a fruitfly's survival.

ADH, it seems, is a versatile enzyme that has evolved in different times and settings. In fact, ADH activity is carried out by three families of enzymes that seem to have arisen independently. The families are spread among most major life forms — from bacteria to plants, yeast and animals. It seems as though the structure of ADH, which allows it to bind to alcohol as well as to several other chemicals, made it a useful enzyme under different circumstances.

The original purpose of the ADH now found in humans probably wasn't breaking down alcohol: the fact that the enzyme can do this simply came in handy later on. So, what was its original function? At the moment, nobody knows. But some are hazarding a guess. Ricard Albalat, an evolutionary geneticist at the University of Barcelona in Spain, believes it was used to break down other potentially harmful chemicals, such as formaldehyde⁴. "Formaldehyde can react with DNA and cause mutations," notes Jan-Olov Höög, a medical biochemist at the Karolinska Institute in Stockholm, Sweden. "The ability to break it down is a crucial function of ADH."

But whatever their true origins, there is clearly a lot more to these multitasking enzymes than just allowing us to get drunk. As researchers delve further into their history, these molecules are shedding light on the big questions of evolutionary biology. A surefire cause for celebration.

Siëlle Gramser is an intern in Nature's Munich office.

"Yeast 'realized' there was a lifestyle opportunity, which involved making large amounts of alcohol." — Steven Benner

1. Thomson, J. M. et al. *Nature Genet.* **37**, 630–635 (2005).
2. Matzkin, L. M. *Mol. Ecol.* **14**, 2223–2231 (2005).
3. Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W. & Hoffmann, A. A. *Science* **308**, 691–693 (2005).
4. González-Duarte, R. & Albalat, R. *Heredity* **95**, 184–197 (2005).

Saving the agave

A decade ago, the tequila industry was pummelled by plant diseases.

Rex Dalton meets the scientists working to keep the blue agave diverse enough to survive.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

For centuries, artisans working in the adobe haciendas of Mexico's rural valleys have followed tradition to make the powerful spirit tequila. Copying age-old indigenous techniques, they distilled the liquor from sweet juice cooked out of the fat stems of a local succulent, the blue agave (*Agave tequilana* Weber, var. *azul*).

But in recent years, tequila makers have had to bring the latest science to the agricultural process to save both the industry and the culture it supports. Some of the oldest and biggest producers are employing scientists, building high-tech laboratories and funding academic research on the blue agave so that researchers from biochemists to geneticists can scrutinize this little-understood plant.

The shift began nearly a decade ago, when disease and pests wiped out much of Mexico's crop of blue agave. The plants are grown in expansive ranches, as a single agave takes years to reach maturity for harvest. But those huge monocultural crops, planted to slake the worldwide thirst for tequila, are also an ideal place for disease to spread. Tequila was nearly destroyed by its own popularity.

The agave plant grows a rounded stem covered with thick, spiked leaves. The plants are harvested at the age of seven years, when sugar content is at its peak. The leaves are cut off, leaving a 'head' that looks like a huge pineapple. Heads are then cooked for the sweet juice, which is fermented and distilled into liquor.

Until about 25 years ago, tequila was known

mostly as a traditional drink in Mexico, rarely savoured outside the country save by college students and the adventurous. But in the early 1980s, enthusiasm for the beverage blossomed as its better-quality varieties became more widely known — with help from songs such as Jimmy Buffett's classic *Margaritaville*.

Boom time

To meet demand, ranchers industrialized the planting process to produce millions of genetically similar blue agave plants for maximum yields. Plantings of blue agave leapt from 16,000 hectares to nearly 50,000 in less than a decade. But by following this route, plant scientists say, the ranchers sowed thousands of

hectares with plants whose lack of diversity left the crop susceptible to devastation when disease struck. Agave plantations are generally all of the same variety. Farmers usually cut off the flowering stalk to increase the plant's sugar load, which means that the plants aren't cross-pollinated by bats or other animals as they would be normally. Without that mixing, the blue agave crop is nearly genetically uniform, a situation that renders it particularly prone to disease. A single pathogen can rapidly destroy most of an entire crop.

"We told them this was going to happen," says Gary Paul Nabhan, an ethnobotanist at Northern Arizona University, Flagstaff, who with Mexican ethnobotanist Ana Valenzuela



Harvest time: the leaves are cut from the head of a mature agave. On average the plant takes seven years to reach this stage.

Zapata wrote a seminal book on the problem¹. "But they wouldn't listen to us."

Beginning in the late 1980s, disease began to rot agaves in the fields. And then, in 1996 and 1997, a climate shift enveloped the prime agave-growing states near Mexico's southwestern Pacific coast. The warmer temperatures and increased rainfall proved devastating for the plants². Diseases and weevils attacked: a bacterium (*Erwinia carotovora*) and a fungus (*Fusarium oxysporum*) were particularly malign, ruining the valuable heads. Ranchers, who must tend fields for years before getting the money from a harvest, were left with unusable agaves. Many decided to cut their losses and abandon agave harvest: by some estimates, the area of planted fields plunged more than 25%.

Producers began to scramble for agaves, and tequila prices skyrocketed. For the first time, tequila makers began tapping other plants for the sugar juices needed for fermentation. Tequila was no longer necessarily made from 100% blue agave — a sacrilege for traditional ranchers and tequila artisans. At least one vocal producer, who spoke out against reducing the tequila standards, was assassinated during public demonstrations.

Traditional tipple

By law, tequila production is limited to five states, with most activity in the state of Jalisco. The Mexican Tequila Regulatory Council certifies two types of tequila: the traditional, which is labelled "100% de agave", and a lesser variety, which can be described as 'tequila' only but must still be made with at least 51% blue agave.

In the aftermath of the crop plagues, tequila producers have turned to scientists to revive their crops. But there are only a few specialists, and a very limited literature in agave science. "You don't find many agave publications in journals," says Eulogio Pimienta Barrios, an ecological physiologist at the University of Guadalajara. Historically, producers held any specialized agave knowledge close for competitive purposes. But that situation is changing, as producers bring in academic researchers for studies. Just two years ago, the chemical structure of the sugar of the blue agave, a fructan, was described for the first time — showing that it holds promise for food products for diabetics³. And last summer, two of the biggest tequila distillers, Herradura and Sauza, signed a cooperative research agreement to share information about agave science.

Villages such as Amatitán, 40 kilometres west of Guadalajara, are the nexus of agave culture, where friends sing songs about tequila at informal gatherings. Here, the family-owned Herradura distillery still produces tequila on its 135-year-old hacienda. The agave heads are cooked in old-style ovens and juices fermented in open-top vats: not very different from the way tequila was made in the nineteenth century. But now providing scientific insight is Aideé Orozco Hernández, a bio-

chemist who has worked for the company for about four years. Orozco has installed a laboratory with sophisticated equipment for research into plant breeding and production techniques.

Another 15 kilometres west, in the town of Tequila itself, sits the distillery of Sauza, owned by a multinational liquor giant. In 1999, the firm hired plant physiologist Ignacio del Real Laborde. The new cooperative scientific approach, says del Real, is a symbol of the need to see beyond local rivalries and think in terms of a global market. "It doesn't help to have your neighbour doing bad things," he says. "Some people weren't doing proper agriculture before, but we are now."

More still needs to be done to adopt sustainable practices, says Valenzuela, who is based at the University of Guadalajara. Ranchers, she says, are reluctant to try new approaches because they fear economic losses.



Agave heads pack more sugar if the plants aren't allowed to flower.

One such issue is whether to allow some plants to flower and so allow cross-pollination. Farmers say they can't afford to lose agaves to pollination and want to maintain their plants' valuable characteristics. "Some are changing," says Valenzuela. "But industry people are not paying enough attention to the erosion of biodiversity."

Orozco and other tequila scientists are trying to address this by creating more diverse agave lines for large-scale planting. Last year, Herradura harvested seeds from experimental fields where the agaves were allowed to pollinate naturally. Some 400 agave lines were selected from this experiment for further study, says Orozco, and the best lines will be chosen for breeding and planting. "We are very confident that new knowledge about the plant will give us more efficiency and quality," she says.

But for del Real, there are limits to applying sustainable agricultural techniques. He doubts

whether natural pollination will ever be widespread in agave cultivation, because he fears that the plants could hybridize, creating a version that wouldn't be considered true agave by the tequila regulators. "Valenzuela's view is very respected," he says, "but we need more basic academic research to understand this plant."

To that end, Sauza is funding projects at half a dozen Mexican universities. At the Center for Research and Advanced Studies in Irapuato, Guanajuato, molecular biologist June Simpson is probing plant genetics. About ten varieties of *A. tequilana* have been identified; her institute's research has shown that all are basically genetically identical, although their colours and shapes can vary slightly⁴.

"When I present these data to agronomists or farmers, they say they can't be true," Simpson says, "because they see differences in the plants." In an added twist, she says, a study now in press has found some diversity among these varieties when sampled from a broader area.

She is working to develop a test to identify blue agave through genetic fingerprinting, and her institute is using genetic markers to explore for certain key genes associated with plant sugar production. "All of this is aimed at understanding real genetic improvement," she says. "Then people can do plant breeding."

Glut and disease

But all these new studies cannot staunch the rising fear that another agave crisis could occur. Mexican officials now are predicting a glut of agaves for at least the next three years, with production peaking at nearly 1.8 million tonnes in 2008. Production facilities can accommodate only about half of those, says Alvaro García Chávez, a rural development official in Jalisco.

With agave prices plunging as a result, many ranchers have cut back on caring for their plants — which, in turn, creates bastions for disease and pests. Already, some estimate that 10% of the agave fields are afflicted with disease. "I hope for the best," says del Real. "But yes, I am concerned."

To create new markets for agaves, government officials have encouraged the development of a diabetic-friendly food syrup based on the agave's fructan. But the project uses only a small fraction of excess agaves, leaving scientists and authorities looking for answers to the plant's boom-and-bust cycles.

In the end, they may have to hope for a renewed thirst for their drink. Perhaps the new US hit country song, Joe Nichols's *Tequila Makes Her Clothes Fall Off*, will again pump up demand.

Rex Dalton is Nature's West Coast correspondent.

1. Valenzuela-Zapata, A. G. & Nabhan, G. P. *iTequila! A Natural and Cultural History* (Univ. Arizona Press, Tucson, 2003).
2. Nobel, P. S., Castañeda, M., North, G., Pimienta-Barrios, E. & Ruiz, A. J. *Arid Env.* **39**, 1-9 (1998).
3. Lopez, M. G. et al. *J. Agric. Food Chem.* **51**, 7835-7840 (2003).
4. Gil Vega, K. et al. *Euphytica* **119**, 335-341 (2001).

R. DALTON

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The grapes of rock

Winemakers in the United States are increasingly calling on the services of geologists to help refine their products. **Alexandra Witze** meets the scientists who are treading a path to the past.

There's more to it than just the taste of tannins, the hint of blackberry, the overlay of toasted oak — and the gentle enticement to intoxication. "Every time you have a glass of wine you're drinking 100 million years of Earth history," says David Howell, a geologist at Stanford University in California.

Consider, for instance, a glass of fine wine from California's Napa Valley. Its taste depends on the grapes from which it is made, the water and climate encountered by the vines that bore those grapes, their pruning and harvesting by the field workers and the craft of the vintner. But it also draws on the fertile alluvial soil that spreads in fans down from the hills — hills that are themselves made of ancient oceanic crust, the remains of a collision between tectonic plates.

Of the many views one can take of a glass of wine, geology looks furthest back in time. And in recent years a number of geologists have turned this view into a way of offering professional services to the wine business, from helping select the best sites for planting to providing remote-sensing imagery of growing grapes. Next March, many of this new breed will gather at the University of California, Davis, for a leisurely three-day conference on the science of wine — followed, naturally, by two days of field trips through Napa and Sonoma county vine-

yards. Tastings are practically mandatory.

The focus of these geological considerations, *terroir*, is somewhat fuzzy. At its simplest, it is the combination of physical factors — soil, climate, environment — that help shape a wine's taste. At its most complex, *terroir* is an interplay involving cultural preferences and a long history of working the land. It can be applied to other products, such as cheeses, that come from a particular and distinctive landscape. It is a concept that can shade into mysticism, or cynicism. Emphasizing the characteristics of a wine's place of origin rather than the grape variety, as the French do, can hint at a unique geological attribute that might be seen as justifying a premium price.

Going deep

"There's a fair amount of black magic involved," says Kenneth Verosub, a palaeomagnetism expert at the University of California, Davis, and organizer of the March conference. And that is where he thinks geology can help. Howell likes to see the concept of *terroir* as a means to an end. "It's a way of letting people know that there's more to wine than just the grapes and the roots and the soil," he says.

Soil has always been a major part of wineries' worries. Soil scientists advise on the best places to plant vines, and hydrologists suggest how best to water and nurture them. But intro-

ducing pure geology — reaching down to the bedrock itself — is a relatively new phenomenon for US wineries.

Often the geologists help by mapping contacts between different geological units, each of which has its own characteristics for growing grapes. Or they might help winegrowers to understand the three-dimensional picture of a vineyard: vine roots can penetrate many metres down, potentially tapping a deeply buried soil type that differs from that at the surface. That is especially important in the United States, where many wineries are set atop thick alluvial deposits on valley floors, unlike the traditional hillside plantings of European vineyards.

And in general, the winemakers seem happy to get scientific advice. "They are extraordinarily interested in learning as much as they can about the land in which they grow," says Jonathan Swinchatt, a Connecticut geologist who has collaborated with Howell.

One example is Warren Winiarski, owner of Stag's Leap Wine Cellars in Napa Valley. Winiarski says that he appreciates the insights of science without feeling that he needs to understand every last equation. "An athlete doesn't have to know physiology in order to run the race," he notes.

Winiarski called in geologists because he wanted to understand why wine made with

grapes grown in neighbouring vineyards tasted so dramatically different. Two of his vineyards, Fay and Stag's Leap itself, sit side by side and yet identical vines planted in each yielded very different wines. Investigating, Swinchatt and Howell traced the difference back to geology: Fay and Stag's Leap rested on separate alluvial fans spilling down from the mountains. With that information, Winiarski has modified his vine selection and growing practices for each plot of land.

Geological studies such as this first became well-known in Napa after the aphid-like insect phylloxera, scourge of vineyards worldwide, began to devastate California grapes in the mid-1980s. The massive die-off caused many winemakers to re-evaluate their land and what should be planted on it. Before that, says Swinchatt, "nobody in Napa would talk about *terroir*. If you talked about *terroir* it would be like giving the French some recognition."

But now *terroir* is becoming a West Coast buzzword. Wine marketers see it as a way to individualize each product and stamp it with regional distinction. "They really appreciate that their geological history makes their vineyards different from vineyards in Kansas or in Italy," says Larry Meinert, a geologist and wine consultant at Smith College in Northampton, Massachusetts.

Ground control

Geology is even making it on to labels and into the names of the wines themselves. Columbia Crest, the largest winery in the state of Washington, has introduced a new brand called Torrent: its back label describes the ice-age floods that poured through the scablands of western Washington, scouring the landscape and fashioning the gravels in which the grapes are now grown. Another Washington winemaker has dubbed his land Loess Vineyard, after the air-deposited soils of the region.

For Alan Busacca of Washington State University in Pullman, this marks a welcome new accuracy in such matters: he is tired of seeing wine labels that tout the 'rich volcanic soils' of the Pacific Northwest. "The volcanic stuff is actually a trivial fraction," he grumbles. "Most of the soils in the northwest are formed from outburst flood deposits, with a mantle of reworked glacial material."

Terroir as a marketing tool is also catching on outside the United States. Five large new wineries in Patagonia that sit on gravels washed down from the Andes plan to market their geological and geographical characteristics aggressively, says Meinert. In New Zealand, the Gimblett Gravels appellation is defined by a single stratigraphic unit: to bear the Gimblett label, at least 95% of the grapes must be grown on that particular kind of gravel. "That's a very satisfying thing," says Meinert.

But sometimes, the geology references can degenerate into a deluge of transferred epithets. "I've heard a winemaker correlating the explosive taste of his wine with the explosive



Ah, the shale: wine-lovers in the film *Sideways* enjoy bottled geology in California's Santa Ynez valley.

nature of the rocks in which it was grown," says Swinchatt, "or talk about red spicy flavours reflecting the red soil." A different form of literalism has inspired Randall Grahm, winemaker at California's renowned and eccentric Bonny Doon winery, to experiment with putting smashed-up rocks into wine as it ages.

To some, such attempts are an unwelcome geological reductionism. "The most important thing about this idea of *terroir* comes from learning how to grow grapes and make wines in particular environments," says Warren Moran, a geographer at the University of Auckland in New Zealand. "Every region where wine is grown has interesting stories

about the way the region developed. That's much more interesting and powerful than any simple environmental determinism." In some locations, traditional viticultural appellations — the formalized descriptions of particular grape-growing regions — happily ignore local stratigraphy. To Moran and other geographers, *terroir* encapsulates far more of the notion of territory than it does of geology.

But the geologists are sure they have something to offer — and, they are happy to admit, something to gain. "I work for wine," says Terry Wright, a retired field geologist and vineyard consultant in Sonoma county. Consulting fees regularly include a bottle of wine or cases at employee discounts. A few scientists have got so deeply into wine that they have made a second career of it: Busacca, for instance, will leave his university in a few months to start a vineyard consulting business.

For geologists, soil scientists and hydrologists looking to get into the field, Meinert advises that they start by learning how sites are chosen for various grapes. "Drink lots of wine and pay attention to what's going on physically," he says.

And for those who do not want to work in wine, but just enjoy it, geology can also help, says Swinchatt. "If you know the provenance of a painting, it can mean a lot more than if you didn't know anything about its history," he says. "I think the same is true of wine. If you know where it came from, how it was produced, what kind of care went into making it, it makes a huge difference to how you appreciate it."

Alexandra Witze is a senior news and features editor in Nature's Washington office.



The roots of the matter: Alan Busacca examining the soil in a Washington state vineyard.

BUSINESS

Merck opts for shake-up to clear drug pipeline

The failure of the painkiller Vioxx and a lack of new products leaves the world's third-largest drug company in the lurch. **Emma Marris reports.**

Substance P is an intriguing neurotransmitter molecule that has frustrated a generation of pharmacologists. For decades, it has been known to be highly active in both the brain and the gut — but it cannot easily be lured into the medicine cabinet.

Researchers at drug giant Merck, in particular, have moved heaven and Earth to study the molecule and associated compounds. But the best they have managed so far is approval for the related drug Emend (aprepitant), which reduces nausea in chemotherapy patients and managed sales last year of only US\$80 million or so. Last month, Merck gave up the chase, announcing that it would close down the research unit at Terlings Park, near London, that had been spearheading its investigation of Substance P.

The pull-back is part of a major restructuring at the world's third-largest drug company, whose reputation has been battered over the past year by the fall-out from the withdrawal of its blockbuster painkiller drug, Vioxx, and by shareholders' concerns over the state of its future drug pipeline (see graph). Some 7,000 employees will lose their jobs as the corporation, headquartered in Whitehouse Station, New Jersey, refocuses its research and streamlines its manufacturing operations.

Rebuilding faith

Late last month, the company outlined its plan, saying that, as well as closing Terlings Park, it would shut down five factories and two labs that do preclinical drug testing.

The company will refocus its research in nine areas of interest: Alzheimer's, atherosclerosis (blocked arteries), heart disease, diabetes, obesity, cancer, pain, sleep and new vaccines. Many diseases in these areas are chronic, so patients buy their medicines for years, providing steady income. The company also claims that it will cut nine months from the time it takes to put a new drug through the late stages of development.

"Merck will remain a research-driven pharmaceutical company," says Richard Clark, a 59-year-old Merck veteran who took over as the company's chief executive in May. "But we

need to change our approach to virtually every aspect of our business, and we must act with a sense of urgency."

The research consolidation is being watched carefully by the rest of the pharmaceutical sector, where Merck has traditionally been noted for a strong commitment to making drug discoveries at its own labs. According to Albert Rauch, an industry analyst at brokerage A. G. Edwards, based in St Louis, Missouri, the company has sometimes held off from acquiring drug companies that don't fit its image as a firm that finds its own drug candidates. "When they start cutting research and development I think you'll see a very large negative reaction from Wall Street," predicts Rauch.

The restructuring should save the company \$1 billion a year between now and 2010. Clark hopes that it will help to restore shareholders' faith in the company to where it was last October, before Vioxx's adverse side effects turned the painkiller from a \$2.5-billion-a-year cash cow into a huge potential liability. Thousands of Vioxx patients are set to sue Merck: of the three cases to go to court in the United States so far, the first was lost by the drug firm, the second won, and the third was declared a mistrial last week. The drug is thought to increase patients' risk of having a heart attack, and fresh revelations about when Merck scientists first learned this — in a *New England Journal of Medicine* editorial — have also knocked the company's share price (see *Nature* 438, 899; 2005).

But while Vioxx grabs the headlines, the humdrum matter of patent expiry on existing drugs is of equally pressing concern to share-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

holders. The recipe for Zocor (simvastatin), a cholesterol-lowering drug, will cease to be exclusively controlled by Merck in the United States next June, when generic drug firms are expected to start supplying it cheaply. In 2004, Zocor accounted for almost a quarter of Merck's \$23 billion in total sales. Even if some patients and doctors stick to the familiar brand name, Merck expects sales to plummet.

Slim prospects

The company's product pipeline does not contain any obvious blockbusters to fill the gap. Some of its more promising candidates — including Substance P as an antidepressant — have fallen by the wayside in the past few years.

And hopes have been dashed for the diabetes drug candidate Pargluva (muraglitazar), which Merck was developing jointly with New York-based Bristol-Myers Squibb. The US Food and Drug Administration recently demanded more tests to check for long-term heart risks, and these will eat up years of the drug's allotted patent protection period. Merck now plans to terminate its partnership with Bristol-Myers Squibb.

Merck has three important vaccines in late-stage development, for infant gastroenteritis, shingles and human papillomavirus — a virus associated with cervical cancer. Clark has said that successful launches of these will be "a key factor" in Merck's future performance. Two new cholesterol-adjusting drugs are also on the drugmaker's horizon, and could be ready for approval by 2007.

Uncertainty about drug prospects is not unique to Merck, of course. "A number of the major companies are seeing similar problems with drugs going off patent," notes Stephan

MERCK SHARE PRICE



IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Vicious circle: Merck is having to close factories and labs, but it badly needs new products.

AP/D. HULSHIZER

Gauldie, a senior analyst at Wood Mackenzie, an Edinburgh-based consultancy. "The likes of Bristol-Myers Squibb and Pfizer are facing a huge threat from generics."

IMS Health, a Connecticut-based pharmaceutical consultancy, says the five biggest drugs to go off patent next year are Zocor, Pfizer's antidepressant Zoloft, Bristol-Myers Squibb's cholesterol-lowering drug Pravachol, Sanofi Aventis's sleeping pill branded as Stilnox or Ambien, and GlaxoSmithKline's Zofran, which prevents vomiting.

The end of patents for these lucrative products has weakened pharmaceutical stocks and led to rumblings that the era of the blockbuster drug may be coming to an end. Some analysts think that most of the best drugs for very common ailments may already have been discovered. On this assessment, the future of the industry lies in the trickier business of niche marketing relatively expensive treatments for rare or complex conditions.

In such an environment, Rauch describes the outlook for Merck as "very gloomy", and says that the company is "trading off its dividends". At present, Merck pays a generous 5% dividend to shareholders, making its stock attractive even to investors who do not expect its price to go up. Clark has already promised that he will keep the dividend at its current level.

Yet observers say that Merck has to offer investors more than a share in the cash bounty of its past proceeds. Unless the drug pipeline is fixed, they say, the company's long-term prospects are bleak — with serious ramifications for its rivals, large and small. "Merck is good for the pharmaceutical industry," says Rauch. "You can't make me-too drugs unless you have someone to copy." ■

IN BRIEF

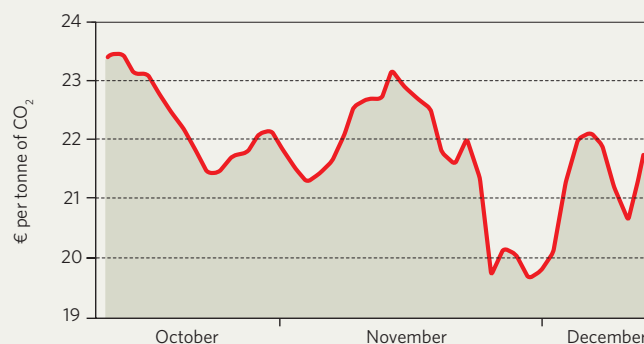
BREATHE EASIER Tobacco giant Philip Morris has embarked on a business alliance aimed at helping premature babies to breathe. The company, based in Richmond, Virginia, has joined up with Discovery Laboratories of Warrington, Pennsylvania, which makes an artificial surfactant — a protein-lipid substance produced in the lungs that is critical for breathing, but often missing from babies who are born more than a month premature. Philip Morris will use its proprietary aerosol technology to develop a device that will deliver the surfactant deep into the lungs.

VIRGIN TERRITORY Plans have been unveiled for a US\$200-million private spaceport to be built near Las Cruces, New Mexico, in 2007. The state's economic development office says it has agreed tenancy terms for the project with Richard Branson's Virgin Galactic, which intends to put its headquarters there and use the facility as a launch site for space tourists. Virgin says 100 people have already paid \$200,000 a ticket for suborbital jaunts on a vehicle to be built by California-based Scaled Composites, which won last year's X Prize for sending a privately developed vehicle to the edge of space.

BIOTECH FIGHTS BACK The US Biotechnology Industry Organization is spearheading a drive to shield small businesses from the requirements of a corporate ethics law that it says is too cumbersome for its member companies. Jim Greenwood, the former Pennsylvania congressman who is now the organization's president, is leading the push to relax the Sarbanes-Oxley law. The 2002 law tightened corporate accountability in response to accounting scandals surrounding the energy conglomerate Enron and other US corporations. Greenwood backed the bill when in Congress.

MARKET WATCH

EEX EMISSIONS ALLOWANCES



The newspapers got excited when nations meeting in Montreal this month agreed plans to negotiate a successor to the Kyoto Protocol. But Europe's nascent market in carbon dioxide emissions took the talks in its stride, barely fluttering in response to the last-minute deal.

After a year of trading on the Leipzig-based European Energy Exchange (EEX), one of five such markets operating in Europe, the price of an allowance to emit one extra tonne of carbon dioxide during 2005–07 seems to be stabilizing at value of about €20 (US\$24). A binding international agreement to cut emissions after 2012 would provide extra security for banks, companies and investors who own the options, analysts say — but improved prospects for the agreement haven't notably increased demand for them.

During the past two months, the price

has stuck in a band between €19 and €24, following much wider fluctuations earlier in the year. Volume continues to grow: in the five markets, between 6 million and 10 million options are now sold every week, up from around 5 million a week in September.

Recent swings have reflected other climate-policy developments, says Marcel Hanakam of Frankfurt-based Climate Change Consulting. Reports in late September that the European Union might make aviation subject to emissions trading boosted the price, for example, and a court ruling that questioned the UK emissions allocation plan caused it to fall earlier this month.

Large companies in the European Union are allowed to emit a certain amount of carbon dioxide a year; if they require additional allowances they need to buy them on the emissions markets. ■

Quirin Schiermeier

For quiet students, finding a voice is the first step towards taking a stand

SIR — As a Chinese graduate student studying in Canada, I often hear stories that reflect your News Feature “Taking a stand” (*Nature* **438**, 278–279; 2005), about my Asian colleagues feeling mistreated by their lab-mates.

The Ontario Human Rights Code states that “individuals have the right to equal opportunities in the workplace and to an educational environment free of harassment because of colour, age, sex, sexual orientation, ethnic origin, religion and handicap”. The United States has a similar law. I would like to think that mistreatment of the Chinese workforce in North American graduate schools is minimal at most, yet I fear that this is not a realistic hope.

Although I currently work in a friendly and cooperative lab and feel fortunate to have helpful and supportive co-workers, my comfort does not mean that discrimination does not exist elsewhere.

A number of my Chinese friends in North America, including one senior postdoc, have, in my opinion, been discriminated against. Unfortunately, a traditional Chinese upbringing encourages passive and nonverbal avoidance of conflict. Furthermore, competition in lab research is fierce, particularly in the United States, so anyone with an obvious weakness such as a language barrier or cultural difference is more likely to be taken advantage of.

To fight for equality in the workplace, one needs to be socially adaptable, and must voice concern if equal rights are being violated.

Peter Cheung

University of Western Ontario,
London Regional Cancer Program,
Room A4-805, 790 Commissioners Road East,
London, Ontario N6A 4L6, Canada

Animal culture is real but needs to be clearly defined

SIR — William Abler, in Correspondence (“Evidence of group learning does not add up to culture” *Nature* **438**, 422; 2005), takes issue with Jacqueline Zupp’s assertion, also in Correspondence (“Concern at animal research should not be dismissed” *Nature* **437**, 1089; 2005), that “we now have evidence for animal cultures”.

Two points relevant to Abler’s concern deserve emphasis in relation to the now-extensive literature on animal traditions, in which terms such as ‘culture’, ‘cultural transmission’ and ‘cultural evolution’ are

routinely applied not only to primates but also to taxa as diverse as birds and fish — references to the ‘cultural transmission of birdsong’, for example, have been familiar for decades.

First, many biologists treat ‘culture’ as a synonym for ‘tradition’, a term defined as objectively as any in the physical sciences. This is no more “emotional vocabulary”, to use Abler’s description, than other everyday terms such as ‘intelligence’, ‘memory’ and ‘innovation’. These terms can also refer to distinctive phenomena in humans, but once objectively defined they are commonly and usefully applied in the science of animal behaviour.

Second, as I explained at greater length in my recent Progress article “The second inheritance system of chimpanzees and humans” (*Nature* **437**, 52–55; 2005), some behavioural scientists do argue that the term and concept of ‘culture’ should be reserved for traditions that share certain sophisticated features with the human case, such as transmission by teaching.

Either perspective can be effectively employed in comparative and evolutionary analyses, but whichever approach is used, the needs of good science remain the same: when we use everyday words such as ‘culture’, they must be clearly defined.

Andrew Whiten

Centre for Social Learning and Cognitive Evolution and Scottish Primate Research Group,
School of Psychology,
University of St Andrews, South Street,
St Andrews KY16 9JP, Scotland

Women's efforts are more than a drop in the ocean

SIR — As an oceanographer I enjoyed your recent Insight on Bio-oceanography (*Nature* **437**, 335–368; 2005). But the only female author in this section was the senior editor who wrote the introduction. This is quite surprising, considering the number of highly qualified women in biological oceanography. In fact, 42% of the members of the American Society of Limnology and Oceanography who are registered as biological oceanographers are women.

I have found that, during the years 2004–2005 (volumes 421–437), *Nature* published 11 Insights with 68 individual overviews, reviews and/or commentaries. Only 10 of the 134 authors were women, and a woman was the first author in only about 4% of cases.

In all cases when the author’s name caused the slightest doubt about their sex, I searched the web for confirmation. In nearly every case there was a picture available — and although I cannot rule out cross-dressing, I have no reason to suspect it is widespread. In the

only case in which I could not find the information I was seeking, I used a name finder to tell me the author’s sex.

A balanced sex ratio is impossible to maintain in peer-reviewed publications, and indeed should not be a goal in peer-reviewed articles. But Insights are written by invitation only, so the editors can decide who they believe could best contribute. The editors should make more effort to promote equality in the process of publication.

Daniel Conley

Department of Marine Ecology,
National Environmental Research Institute,
DK-4000 Roskilde, Denmark, and
Department of Marine Ecology, Aarhus
University, DK-8200 Aarhus, Denmark

Network aims to make maths count in Africa

SIR — Your Editorial “Networks for Africa” (*Nature* **438**, 395; 2005) raises the question of how research and teaching in the mathematical and physical sciences in Africa can best be strengthened. A vital ingredient is surely that, with whatever assistance richer nations can provide, the broader scientific community across Africa should itself plan and manage its own scientific development.

It is in this spirit that the African Academy of Sciences and the International Mathematical Union support a distributed network of African mathematicians in the African Mathematics Millennium Science Initiative, or AMMSI (www.ammsi-maths.org). AMMSI supports research and postgraduate training in mathematics at universities in sub-Saharan Africa. Individual grants are awarded to students and faculty members whose low salaries, high teaching loads and geographic isolation have inhibited their full functioning as teachers, mentors and researchers.

Since 2004, AMMSI and the African Institute for Mathematical Sciences in Cape Town have been working to ensure that the African mathematical community is a fully vested partner in the proposed IT infrastructure network mentioned in your Editorial: the African Mathematical Institutes Net, or AMI-Net.

John Ball

International Mathematical Union,
Mathematical Institute, 24–29 St Giles’,
Oxford OX1 3LB, UK

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Correspondence letters must be signed by no more than three authors; preferably by one. Published contributions are edited.

COMMENTARY

Barriers to progress in systems biology

For the past half-century, biologists have been uncovering details of countless molecular events. Linking these data to dynamic models requires new software and data standards, argue **Marvin Cassman** and his colleagues.

The field of systems biology is lurching forwards, propelled by a mixture of faith, hope and even charity. But if it is to become a true discipline, several problems with core infrastructure (data and software) need to be addressed. In our view, they are too critical to be left to *ad hoc* developments by individual laboratories.

Systems biology has been defined in many ways, but has at its root the use of modelling and simulation, combined with experiment, to explore network behaviour in biological systems — in particular their dynamic nature. The need to integrate the profusion of molecular data into a systems approach has stimulated growth in this area over the past five or so years, as worldwide investments in the field have increased. However, this early enthusiasm will need to overcome several barriers to development.

A recent survey carried out by these authors — conducted by the World Technology Evaluation Center (WTEC) in Baltimore, Maryland, and funded by seven US agencies — compared the activities of system biologists in the United States, Europe and Japan¹. The survey reveals that work on quantitative or predictive mathematical modelling that is truly integrated with experimentation is only just beginning. Progress is limited, therefore, and major contributions to biological understanding are few. The survey concludes that the absence of a suitable infrastructure for systems biology, particularly for data and software standardization, is a major impediment to further progress.

Come together

The WTEC survey confirmed that vital software is being developed at many locations worldwide. But these endeavours are highly localized, resulting in duplicated goals and approaches. Tellingly, one Japanese group called their software YAGNS, for 'yet another gene network simulator'. There are many reasons for this cottage industry: the need to accommodate local data; the requirements of collaborators to visualize data; and limited knowledge of what is already available. In

general, however, it is a terrible waste of time, money and effort. Most software remains inaccessible to external users, even when the developers are willing to release it, because supporting documentation is so poor.

For software developers and skilled users these problems are not insurmountable. But sharing of the benefits of systems biology more widely will occur only when working biologists, who are not themselves trained to develop and modify such software, can manipulate and use these techniques. Unfortunately, the translation of systems biology into a broader approach is complicated by the innumeracy of many biologists. Some modicum of mathematical training will be required, reversing the trend of the past 30 years, during which biology has

become a discipline for people who want to do science without learning mathematics.

A reasonable set of expectations is that different pieces of shared software should work together seamlessly, be transparent to the user, and be

sufficiently documented so that they can be modified to suit different circumstances. Funding agencies would be unwise to support software development without also investing in the infrastructure needed to preserve and enhance the results. One way to do this would be to create a central organization that would serve both as a software repository and as a mechanism for validating and documenting each program, including standardizing of the data input/output formats.

As with centralized databases, having a shared resource with appropriate software-engineering standards should encourage users to reconfigure the most useful tools for increasingly sophisticated analysis. A group sponsored by the US Defense Advanced Research Projects Agency, and involving one of us (M.C.), has developed a proposal for such a resource². This repository would serve as a central coordinator to help develop uniform standards, to direct users to appropriate online resources, and to identify — through user feedback — problems with the software. The repository should be organized through consultation with the

community, and will require the support of an international consortium of funding agencies.

Diverse data

The problems with software diversity are mirrored by the diversity of ways that data are collected, annotated and stored. Such issues are even worse than those faced by the DNA-sequencing community, because experimental data in systems biology is highly context dependent. For data to be useful outside the laboratory in which they were generated, they must be standardized, presented using a uniform and systematic vocabulary, and annotated so that the specific cell type, growing conditions and measurements made — from metabolite- and messenger-RNA-profiling to kinetics and thermodynamics — are reproducible.

Easy access to data and software is not a luxury, it is essential when results undergo peer review and publication. For the scientific community to evaluate the increasingly complex data types, the increasingly sophisticated analysis tools, and the increasingly incomplete papers (that cannot include all information because of the very complexity of the experiments and tools), it is vital that it has access to the source data and methods used.

Dealing with these complex infrastructure issues will require a focused effort by researchers and funding agencies. We propose that the annual International Conferences on Systems Biology would be an appropriate venue for initial discussions. Whatever the occasion, it must be done soon. ■

Marvin Cassman lives in San Francisco, California, USA.

Co-authors are Adam Arkin of the Bioengineering Department, University of California, Berkeley; Fumiaki Katagiri of the Department of Plant Biology, University of Minnesota, St Paul; Douglas Lauffenburger of the Biological Engineering Division, Massachusetts Institute of Technology, Cambridge; Frank J. Doyle III of the Department of Chemical Engineering, University of California, Santa Barbara; and Cynthia L. Stokes who is at Entelos, Foster City, California.

1. Cassman, M. et al. *Assessment of International Research and Development in Systems Biology* (Springer, in the press) www.wtec.org/sysbio
2. Cassman, M., Szitapanovits, J., Lincoln, P. & Shastry, S. S. *Proposal for a Software Infrastructure in Systems Biology* www.csl.sri.com/users/lincoln/SystemsBiology/Sl.doc

"During the past 30 years biology has become a discipline for people who want to do science without learning mathematics."

BOOKS & ARTS

Pulling the strings

Mathematics holds the key to a unified theory of the Universe.

Hiding in the Mirror: The Mysterious Allure of Extra Dimensions, from Plato to String Theory and Beyond

by Lawrence M. Krauss

Viking: 2005. 288 pp. \$24.95

Michael Atiyah

The search for the fundamental physical laws that govern the Universe has fascinated and driven humans for centuries. Its modern form was essentially launched by Isaac Newton, building on the pioneering work of his predecessors, notably Johannes Kepler and Galileo. From the publication of Newton's *Principia* onwards, it has been a remarkable story, which has delved into every realm of the physical world from the subatomic scale to that of the cosmos and the Big Bang. The theoretical framework that supported this great enterprise has, following the path firmly established by Newton, been based on mathematics. At every major step, physics has required, and frequently stimulated, the introduction of new mathematical tools and concepts. Our present understanding of the laws of physics, with their extreme precision and universality, is only possible in mathematical terms.

This mathematical take-over of physics has its dangers, as it could tempt us into realms of thought which embody mathematical perfection but might be far removed, or even alien to, physical reality. Even at these dizzying heights we must ponder the same deep philosophical questions that troubled both Plato and Immanuel Kant. What is reality? Does it lie in our mind, expressed in mathematical formulae, or is it 'out there'. The recent developments in modern physics that go by the deceptively simple name of 'string theory' bring these age-old questions back to the fore, and are the focus of *Hiding in the Mirror*, a new book by Lawrence Krauss.

Most popular-science books are written by enthusiastic protagonists who seek to convey their excitement to the general public. This book is refreshingly different. On the big questions, Krauss remains a sceptic, a hard-nosed physicist questioning the mountain of mathematical theory that has yet to produce any experimental evidence. But this is no debunking exercise; Krauss makes a serious attempt to bring the reader to the very frontier of modern physics. He describes the intricate theoretical constructs that have been erected

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Out for the count: is physics being taken over by mathematics?

and he acknowledges that, at present, there is no alternative on the table: "it's the only game in town".

One of the book's main themes is indicated by its subtitle, which refers to "the mysterious allure of extra dimensions". This guides the reader through the ever-increasing complexity of physical theory. As Krauss points out, the mystery of extra dimensions exercised the imagination long before it entered serious physics. We are enchanted by the account of Edwin Abbott's *Flatland*, a nineteenth-century classic that combines a scientific and philosophical aim with social satire worthy of Jonathan Swift. The erudite but volatile British mathematician James Joseph Sylvester tried to refer to extra dimensions as 'inconceivable' (by analogy with 'imaginary' numbers). Fortunately for Einstein and subsequent physicists, the term never caught on.

Krauss gives pride of place to Einstein's revolutionary ideas that showed, first in his special theory of relativity and even more convincingly in his subsequent general theory, that space and time form an indissoluble four-dimensional continuum. This was undoubtedly

a landmark in the history of human thought and fully justifies Einstein's iconic status. As Krauss explains, Einstein's general theory, which supplanted Newton's theory of gravitation, was reached by a process of pure thought — not by the pressure of unexplained experimental data. Einstein was motivated by aesthetic considerations, an impressive example of the power of beauty to act as a guiding light. On this point, Krauss quotes the mathematician Hermann Weyl: "My work always tried to unite the true with the beautiful, but when I had to choose one or the other, I usually chose the beautiful."

Krauss comments that mathematicians, poets, writers and artists can all choose beauty over truth, but scientists do not have this luxury.

This is the final word in the book, but Krauss had already mentioned an episode involving Weyl in which beauty eventually triumphed. After Einstein had explained gravity in terms of the curvature of four-dimensional space-time, Weyl attempted to explain electromagnetism in similar terms. Einstein pointed out a fatal flaw in Weyl's explanation (not in the mathematics, but in its physics). Remarkably, Weyl did not withdraw the paper, and the paper was published together with Einstein's objection as an appendix. Clearly the beauty of the mathematics exercised its own appeal.

A few years later, after the appearance of quantum mechanics, a new physical interpretation of Weyl's mathematics was possible; this was the beginning of modern 'gauge theory', the basis of elementary particle physics. To a mathematician this seems a remarkable case of beauty winning through in the end, justifying Weyl's preference. After all, beauty is subjective, it is in the eye of the beholder, and we can be certain about ourselves. Truth is much

more elusive and can change as new facts or ideas emerge.

Krauss's book is well written for a general audience and puts the scientific advances into a historical and philosophical context while keeping the technicalities under control. After a rapid overview of the past it focuses on the current aim of combining the two fundamental theories of twentieth-century physics: Einstein's general theory of relativity, and quantum mechanics, which between them deal with the very large and the very small. The need to unify these two theories is entirely aesthetic; there seems to be little need from the point of view of the experimentalist.

Over the past few decades, string theory has emerged as a serious contender to be such a unified theory. It involves extra dimensions galore: not just Einstein's four, or the five that also incorporates electromagnetism, but a total of 10 or 11. The extra dimensions are viewed as very small and curled up, so that, for most purposes, we are not aware of them. They only manifest themselves at very high

energies, of the kind encountered in particle accelerators. Moreover, these extra dimensions are constrained by very precise symmetry requirements. The upshot of this is that string theorists can exhibit plausible models of a unified Universe, but unfortunately they cannot explain why we inhabit a particular one.

The mathematics involved in string theory is quite remarkable by any standards. In subtlety and sophistication it vastly exceeds previous uses of mathematics in physical theories. Almost every part of contemporary mathematics is involved somewhere in the story. Even more remarkable is that string theory has led to a whole host of amazing results in mathematics in areas that seem far removed from physics. To many this indicates that string theory must be on the right track. But Krauss is not a mathematician, so perhaps he is unaware of all this mathematical success, or maybe he discounts it as irrelevant. Time will tell. ■

Michael Atiyah is president of the Royal Society of Edinburgh, 22-26 George Street, Edinburgh EH2 2PG, UK.

The greatest challenge for Lightman is to give substance to his claim that "the first reports of the great discoveries of science are works of art", an assertion often made but rarely demonstrated. He makes it especially difficult for himself, as instead of giving an overview of the works' literary qualities, he introduces each paper (or sometimes a pair of papers) separately. The result is that the book is a series of 22 essays, each followed by the paper or papers that he has discussed. He reproduces some of them in full, but sensibly cuts the rest of them — a few by as much as two-thirds, others by only a fifth. The research papers are all in English, Lightman having found lucid translations of the seven papers in his selection that were originally written in German.

The physics-related chapters are, predictably, the most accomplished. Best of all is his essay on Hubble's law, which led to the realization that the Universe is expanding. It opens like a novel, on a chilly evening in the late 1920s, the sky "a deep purple gash flecked with stars". Lightman then paints a vivid picture of Edwin Hubble and explains why the discovery made such an impact. The problem is that he does much the same for all the other topics too, so the book is somewhat formulaic and repetitive. As one would expect of this author, the writing is unfailingly clear, but it disappointingly lacks his usual grace and style. Most of his explanations are surprisingly lacking in flair, and are little better than those conventionally served up elsewhere. Nor is the style especially pleasing: I would never have expected this most elegant of science writers, for example, to introduce Max Planck's ideas on quantization to a lay audience by referring to "elemental vibrating resonators".

I had hoped Lightman would persuade me that the finest scientific papers are often great art. Alas, it was not to be. I found myself guiltily flicking through the papers that are outside

Bitesize breakthroughs

The Discoveries: Great Breakthroughs in 20th Century Science Including the Original Papers

by Alan Lightman

Pantheon: 2005. 576 pp. \$32.50

Graham Farmelo

Tapas are one of the greatest pleasures of Spain. These delicious snacks and appetizers are one of the foundations of the country's cuisine. Tapas-style books are becoming common too, as the average attention span of modern readers falls by the year. Here we have a promising science book in the genre, *The Discoveries*, a collection of short pieces on 25 of the best research papers of twentieth-century science.

It is an appealing idea, all the more attractive in this case for being prepared by the much-admired writer Alan Lightman, a physicist and adjunct professor of humanities at Massachusetts Institute of Technology. The author of three well-crafted novels, several popular-science books and many elegantly written essays, he is well qualified to achieve his ambitious aim of providing an insightful overview of modern science.

In his introduction, Lightman says that he sought to find the patterns of discovery, and to compare their discoverers and the different styles of working and thinking among leading scientists. He spent six months consulting widely before he made his final selection of discoveries. In his description at the end of the process, he is winningly open about his passion for science: "I held the stack of twenty-five papers in my arms, a century of scientific thought. My eyes filled with the tears."

Any selection of science's 'greatest hits' is bound to be controversial. But it seems to me that Lightman's choice is reasonable, if rather biased towards physics. He includes papers on quantum theory, Einstein's special (but not general) theory of relativity, nuclear physics, cosmology, Linus Pauling's pioneering paper on the chemical bond, Alexander Fleming's discovery of penicillin, Barbara McClintock's jumping genes, the structures of DNA and haemoglobin, and the first demonstration of genetic engineering. The most striking omission is a paper on plate tectonics, one of the few authentic revolutions of modern science.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Steven Weinberg (right) won the 1979 physics Nobel with Sheldon Glashow (left) and Abdus Salam.

my bailiwick in physics, getting impatient with the technicalities. Lightman's introductions rarely gave me an appetite for unfamiliar fare: a bite or two was quite enough. I suspect that non-physicists will feel the same when they come across the three-line master formula at the heart of Steven Weinberg's unified theory of electromagnetism. They will not, I fear, have

been much encouraged to persevere by Lightman's comment: "Even without knowledge of any of the symbols or their meanings, one must be impressed" by the formula's "economy and power". Some hope.

I have long been an admirer of Lightman, and was expecting *The Discoveries* to be an elegant and palatable introduction to modern

science. Sadly, it is instead an indigestible and tedious read that I believe will have only limited appeal. One of the most creative chefs of science writing has shown that tapas are not his forte.

Graham Farmelo is a senior research fellow at the Science Museum, South Kensington, London SW7 2DD, UK.

A Titan of physics

Huygens: The Man Behind the Principle

by C. D. Andriessse

Cambridge University Press: 2005. 360 pp. £55

Owen Gingerich

Had Isaac Newton never lived, Christiaan Huygens would have iconic status for characterizing physical science in the second half of the seventeenth century. Like Newton, Huygens made enormous contributions in mathematics, mechanics and optics. He anticipated Newton in finding the formula for acceleration in the case of circular motion and brilliantly used it to determine the value of the constant of gravitational acceleration, *g*. He invented the pendulum clock, correctly interpreted the rings of Saturn, found the formula of the catenary curve adopted by a chain fixed at each end, and enunciated the fundamental principle of the wave motion of light.

Huygens was born in Holland in 1629, the second son of a domineering father, Constantijn, who was both a poet and a government diplomat. Christiaan's older brother, also named Constantijn, became a military officer and worked both independently and cooperatively with his younger sibling in making telescope lenses. In 1666, Christiaan, with his reputation as a mathematician already well established, went to Paris to play a leading role in the formation of Louis XIV's new Académie des Sciences. But in 1681, following the death of the minister Jean-Baptiste Colbert, whose patronage had energized the academy, Huygens was no longer welcome in France as the country turned against the Protestants.

In 1661 Huygens had visited London, meeting Robert Boyle and Robert Hooke, where he observed a transit of Mercury across the face of the Sun. In 1689, around the time of William of Orange's coronation as king of England, he again visited London, where he met Newton and Edmond Halley at a meeting of the Royal Society. There was, however, little love lost between Newton and Huygens.

Unlike Newton, who has an abundance of substantial biographies, accounts in English on the life and works of Huygens have been few and far from adequate. This biography by C. D. Andriessse, a physicist at Utrecht University, brings a wealth of newly translated information, making it the richest source of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Christiaan Huygens emerges from Isaac Newton's shadow.

information in English about the seventeenth-century Dutch polymath. The book makes ample use of Huygens' surviving correspondence, diaries and notebooks, as well as his published volumes. Huygens was a somewhat erratic publisher, often holding back works for many years (and thus occasionally losing priority), so having access to the manuscripts was an essential part of this project.

Andriessse's book is a fascinating account, but is by no means an easy read. The flow is interrupted from time to time by technical interludes that explain, for example, Huygens' work with musical temperaments or the production of an isochronous pendulum. These require the reader to be familiar with terms such as 'tonic' or 'evolute'. However, such sections can be easily skipped by a reader impatient with these illuminating mathematical excursions.

More problematic is a torrent of proper

names, both of people and geographical locations. For someone familiar with Dutch history and geography, these may pose no difficulty, but the account would have been rendered more widely accessible with a few strategically placed maps and a glossary of personal names. For instance, the first chapter, which is entitled 'Titan', ends with a paragraph concerning Huygens' discovery of the brightest satellite of Saturn, which he named Titan. Andriessse concludes by remarking that Titan is a fitting image for his subject, quoting a Latin couplet written by Huygens, translated as:

*Let them remain as signs of my
sagacity, and their names
That I write across the heavens
be an echo to my fame.*

Thereafter Andriessse often (and rather confusingly) refers to Huygens simply as Titan.

What makes the book an erratic read are the long sections from letters or diaries, filled with trivia (albeit colourful) and innuendo (regarding

attractive ladies whom Huygens may or may not have taken to bed); these are interspersed with details of his mathematical or scientific achievements. My lingering impression is that the book is too uneven, and even perhaps too disturbing, to be recommended with enthusiasm.

On deeper reflection I realize that the book mirrors Huygens' own personality and psychology. Huygens was beset by painful episodes of melancholy when for many months he seems to have accomplished nothing, followed by great spurts of creative frenzy. The development of the wave theory of light, leading to the principle of the book's subtitle, occurred after a particularly devastating melancholic episode. Andriessse goes so far as to say: "It is thanks to this crisis that we have Christiaan's magnificent piece of work on light." All of this suggests to me that Huygens might well have suffered from

AKG-IMAGES/NIMATALLAH

bipolar disorder (manic-depressive illness).

Huygens was never as interested in philosophy as his contemporaries Newton or Leibnitz, but in his sixties he nevertheless managed to write a more general view of the Universe, his *Cosmotheoros*, and once more his scientific and instrumental genius flashed forth. He devised a way quantitatively to reduce the

brilliance of sunlight to that of the star Sirius, thereby photometrically determining the distance to a typical nearby star. "What bounds of number must we set, especially if we consider the infinite Power of God!" he exclaimed. "Really, when I have been reflecting thus with myself, methought all our Arithmetick was nothing, and we are vers'd but in the very

Rudiments of Numbers." It was his last great work. As the printing began, his health steadily deteriorated, possibly from cancer, and he died before the book was published, in 1695. ■

Owen Gingerich is a historian of astronomy at the Harvard-Smithsonian Center for Astrophysics and author of *The Book Nobody Read: Chasing the Revolutions of Nicolaus Copernicus*.

A vision of birth

A nativity scene painted by Hugo van der Goes bears a medical message.



Martin Kemp

Christmas inevitably brings with it traditional images of the nativity of Jesus Christ. Many show the Virgin Mary kneeling before her son, who lies naked on the ground. We tend to accept this imagery without a second thought, because it is so familiar. But it arose at a particular point in history and carried with it specific associations and meanings that could be adapted to specific contexts.

The image of the Virgin Mary kneeling comes from one of the visions of Saint Bridget, a fourteenth-century Swedish noblewoman. Her vision, she said, made her an eye-witness to Christ's birth: "The Virgin, kneeling with great reverence, placed herself in prayer, with her back to the crib. And while she thus remained at prayer, I beheld her child move in her womb, at once in a moment and in a twinkling of an eye, she brought forth her son...I could not perceive how... she brought forth...the glorious babe lying naked and most pure on the ground."

The idea of a birth that was miraculously quick and painless served to reinforce the dogma of the virgin birth and the doctrine of the Immaculate Conception. Mary was free of the sins and stains that women suffered following the fall in the Garden of Eden.

In one of the greatest of all paintings of Saint Bridget's account of the nativity, this message was adapted for a particular medical context. Hugo van der Goes' huge three-panel altarpiece was commissioned in Bruges by a banker for the Medicis, Tommaso Portinari, and his wife Maria in about 1475. It was shipped to Portinari's native Florence on its completion a few years later. The central panel depicts the nativity with the shepherds, Joseph, angels in ecclesiastical garments, and the ubiquitous ox and ass. The left panel contains Tommaso with two sons and two male saints; in the one on the right, Maria is accompanied by one daughter and two female saints.

This great painting was destined for the chapel of Sant'Egidio, which was attached to the Hospital of Santa Maria Nuova. With 220 or so beds arranged in men's and women's wards, and a staff of physicians, surgeons and apothecaries, the hospital served as a European model in its emphasis on curative procedures.

An obvious medical allusion is apparent in the painting's foreground. The vase containing the irises and lilies is an albarello, almost certainly from Valencia, of the kind used specifically for the storage of medicinal

herbs and minerals. The Venetian drinking glass beside it contains columbines and carnations, which, like the lily, iris and violets scattered on the ground, were used extensively for therapeutic purposes.

Less obviously medical is the miraculous nature of Christ's delivery. However, the presence of the chapel and the emphasis on devotion in effecting cures and alleviating suffering reminds us that the health of the spirit and the well-being of the body were conjoined in Renaissance medical practice. The Virgin Mary, through her painless birth, could act as an inspiration for those in pain to rise above their suffering through spiritual contemplation.

For a twenty-first-century viewer concerned with childbirth, the image may bear other resonances. The favoured birth position in the West from the eighteenth century onwards — lying on the back — has been challenged by those who advocate a return to more traditional and 'natural' methods, including positions that involve kneeling. Perhaps for Saint Bridget, mother of eight children, kneeling to give birth was not that extraordinary, but the absence of pain was undoubtedly unique.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX11PT, UK.

Unravelling string theory

String theory may provide the best clues yet about how to obtain a unified theory that describes all the laws of nature, but do we even understand what string theory is?

Edward Witten

Albert Einstein famously devoted the later part of his life to seeking a theory that would offer, at least in principle, a comprehensive description of the laws of nature. This 'unified field theory', Einstein believed, would endow all of nature's laws with the beauty of general relativity. Ultimately, Einstein left us with plenty of inspiration, but not many ideas about how to proceed.

In fact, there are ample reasons why one might doubt whether Einstein's vision is achievable, or at least achievable in the foreseeable future. Crucial clues may be hopelessly out of reach. When looking back at Einstein's own work, most physicists would say that many of the most important clues for a unified field theory — involving strong and weak nuclear interactions, the role of gauge theory and the world of elementary particles — were simply not known in Einstein's day.

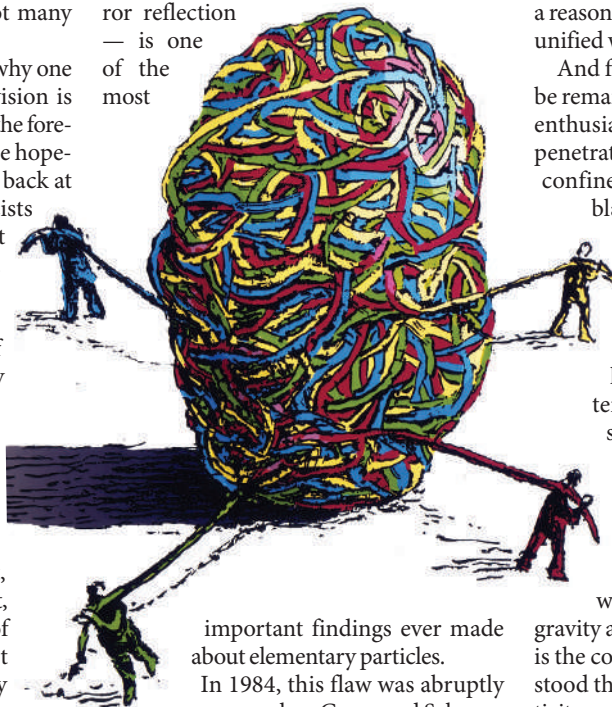
Moreover, even if we could somehow find the unified field theory, it is not at all clear whether we could determine that it is right. From a simple combination of Planck's constant, the speed of light, and Newton's gravitational constant, one can construct a natural unit of length — the Planck length. First defined by Max Planck a century ago, this length is so fantastically small that if it, or something close to it, is fundamental in physics, then some of the most important phenomena may be permanently beyond our experimental reach.

I remember vividly how impressed I was in 1981 when the distinguished experimentalist Norman Ramsey (who later won the Nobel prize for his work on magnetic resonance) forecast that within 50 years there would be a clear outline of a unified field theory, "with all the forces fitting in together, even if not perfectly". I certainly did not see any useful way to work on such a thing, and doubted that I would see it.

Meanwhile, I was only dimly aware of the work that was being done by Michael Green, John Schwarz, Lars Brink — and very few others — to revive string theory. Originally a candidate 'theory of the nuclear force', string theory had been developed and discarded a decade earlier. Its revival was motivated by the hope that it would give the basis for a unified field theory.

By 1982 or 1983, I began to notice this

work, which had advanced to the point that it was possible to formulate fairly convincing theories of quantum gravity unified with matter. There was, however, what seemed like an obvious flaw. String theory seemed incompatible with 'parity violation' in weak interactions (such as nuclear beta decay). Parity violation — the fact that the laws of nature are not invariant under a mirror reflection — is one of the most



important findings ever made about elementary particles.

In 1984, this flaw was abruptly overcome when Green and Schwarz discovered an elegant new mechanism of 'anomaly cancellation'. Not only could the weak interactions violate parity but, especially after the invention of the heterotic string, it soon became possible to derive semi-realistic models of elementary particles with all their known forces, including gravity. At this point, it really did seem reasonable to work on unified field theory.

I suppose that there are three basic reasons why string theory has attracted so much interest in the past 20 years. One is that it is there. String theory is the only known generalization of relativistic quantum field theory that makes sense. The framework of special relativity plus quantum mechanics is so rigid that it practically forces quantum field theory upon us. The tightness of the modern framework is one of the main reasons why physicists were able to discover what has become the standard model of elementary particles. A big idea like a consistent generalization of quantum field theory comes along only

every now and then. So we are duty-bound to take it seriously.

A second reason has to do with what physicists have learned in developing string theory. String theory forces general relativity upon us, whereas standard quantum field theory apparently makes it impossible to incorporate general relativity. And string theory leads in a remarkably simple way to a reasonable rough draft of particle physics unified with gravity.

And finally, string theory has proved to be remarkably rich, more so than even the enthusiasts tend to realize. It has led to penetrating insights on topics from quark confinement to quantum mechanics of

black holes, to numerous problems in pure geometry. All this suggests that string theory is on the right track; otherwise, why would it generate so many unexpected ideas? And where critics have had good ideas, they have tended to be absorbed as part of string theory, whether it was black-hole entropy, the holographic principle of quantum gravity, noncommutative geometry, or twistor theory.

But what is string theory? It may well be the only way to reconcile gravity and quantum mechanics, but what is the core idea behind it? Einstein understood the central concepts of general relativity years before he developed the detailed equations. By contrast, string theory has been discovered in bits and pieces — over a period that has stretched for nearly four decades — without anyone really understanding what is behind it. As a result, every bit that is unearthed comes as a surprise. We still don't know where all these ideas are coming from — or heading to.

One day we may understand what string theory really is. But even if we do, and the theory is on the right track, will we be able to learn how it works in nature? I certainly hope so. Realistically, it all depends on many unknowns, including the nature of the answer, how clever we will be, and the clues we can get from experiment. ■

Edward Witten is at the Institute for Advanced Study, School of Natural Sciences, Princeton, New Jersey 08540, USA.

FURTHER READING

Ramsey, N. F. *Phys. Today* **34**, 26–34 (November 1981).
Zwiebach, B. *A First Course in String Theory* (Cambridge Univ. Press, 2004).
Greene, B. *The Elegant Universe* (W. W. Norton, 1999).

J. KAPUSTA/IMAGES.COM/CORBIS

ESSAY

The death of a star

When Subrahmanyan Chandrasekhar asked in his twenties, 'What happens to a massive star when it runs out of fuel?' he had little idea that it would take a generation of astronomers to find the answer.

Freeman Dyson

Subrahmanyan Chandrasekhar, known to his friends and colleagues as Chandra, opened the door to our understanding of the death of stars. He was the first to calculate the possible final states of stars that have used up their supplies of energy. He did so in 1930, when he was a graduate student travelling by ship from his home in India to study at the University of Cambridge, UK.

Even before he got to Cambridge, Chandrasekhar knew more about relativity and quantum mechanics than most of his teachers. He knew how to take account of both when building mathematical models of cold stars that had stopped shining. On board the ship, once he had finished his calculations, he came to a startling conclusion: he found that there exists a critical mass, now known as the Chandrasekhar limit, beyond which no cold star made of ordinary matter can exist. He calculated this critical mass, and found that it is a few times the mass of the Sun, the exact value depending on the chemical composition of the star.

When a star of less than the critical mass has used up all its fuel, it will slowly radiate away its energy and cool down to reach a state described by one of the models he had calculated. But once a star with greater than the critical mass has used up its fuel, it cannot cool down gradually and die quietly. It must either change into some totally different form of matter, or end its life in a violent collapse and explosion.

When Chandra discovered the critical mass, he had no idea what the ultimate fate of a massive star should be. He opened the door to understanding by raising the question: what happens to a massive star when it runs out of fuel and has no way to cool down?

The efforts of a whole generation of astronomers were needed to find the answer to Chandra's question, starting with Fritz Zwicky's observations of supernovae in the 1930s and ending with the identification of stellar-mass black holes using X-ray telescopes in space in the 1960s. We now know that stars with a mass greater than the Chandrasekhar limit mostly die in catastrophic explosions, which we call supernovae, leaving behind collapsed cores which may be either neutron stars or black holes. Chandra's question led the way to the modern view of the Universe as a dynamic arena dominated by violent events.

When Chandra arrived in Cambridge in



BETTMANN/CORBIS

Star performer: Subrahmanyan Chandrasekhar's insight helped to revolutionize astronomy.

1930, his mentors had no inkling of the revolution that his question was to bring about. Chief among his mentors were Arthur Eddington and Edward Milne, two world-famous astronomers who thought they knew everything worth knowing about stars. Each of them had a private theory of the Universe that was incompatible with Chandra's calculation. They ignored his arguments and declared publicly that his conclusions were wrong.

But Chandra had a cool head. He published his work in reputable astronomical journals and waited for the next generation of astronomers to recognize its importance. He stayed in Cambridge for seven years and remained on friendly terms with Eddington and Milne. After their deaths many years later, he wrote warm and sympathetic memorial lectures for each of them.

Once I went for a long walk with Chandra in the woods around Princeton and listened to him talking about his friendships. His love and admiration for Eddington and Milne were genuine. He saw them clearly, on the one hand as misguided fools, and on the other hand as human beings of rare quality, worthy of honour and respect.

In 1937 Chandra moved to the University of Chicago, where he worked until his death in 1995. His output of research followed a regular pattern. At the beginning of each decade, he chose a fresh field of study. Then he wrote a series of papers solving the outstanding problems in that field. At the end of the decade he published

a magisterial book, summarizing his results and presenting the whole field in a new and clearer light. He worked in each of six fields in turn: in his third decade, he worked on the structure of dying stars; in his fourth on the transport of radiation through stellar atmospheres; in his fifth on instabilities of fluid motions; in his sixth on Einstein's general theory of relativity; in his seventh on the theory of black holes; and in his eighth on a detailed historical study of Newton's *Principia Mathematica*.

Everything that Chandra did was done with elegance and style. He reached a deep understanding of the mathematical and physical properties of black holes, those objects of perfect symmetry that he saw as the crowning beauty of the Universe, a beauty to which Eddington and Milne and even Einstein had been blind. His book about black holes displays his unrivalled mathematical skill as well as his impressive command of the English language.

In his eighth decade, his first great discovery, the Chandrasekhar limit, was recognized with the award of a belated Nobel prize. His last book *Truth and Beauty* is a collection of meditations about the place of beauty in science, including a critical comparison of Newton with Shakespeare and Beethoven, and ending with eloquent tributes to his old enemies Eddington and Milne. ■

Freeman Dyson is at the Institute for Advanced Study, Princeton, New Jersey, USA.

NEWS & VIEWS

MARS

The flow and ebb of water

Mark A. Bullock

Information is pouring in about Mars. These are thrilling times for those who are proposing — and challenging — ideas about the chemical evolution of the planet and its potential for having harboured life.

Was Mars ever wet and warm for long enough to have been a crucible for life? Taking the existence of liquid water as a necessity for life, argument about that question has lately become increasingly intense and fast-moving. The newest proposals appear in two papers in this issue^{1,2}, in which volcanic activity and meteorite impact are respectively put forward to explain martian chemistry previously invoked as evidence for the action of water.

That water once flowed on the surface of Mars seems clear from decades of awe-inspiring spacecraft images of valley networks and giant fluid-carved channels. Most of the valley networks are in the most ancient terrains, however, and were possibly formed only by the melting of ice by impacting debris left over from the formation of the Solar System. The magnificent channels that debouched into the northern plains apparently released as much water as is found in the Mediterranean Sea, but they too were probably ephemeral. And yet, for decades, extensive spectroscopic searches for water-altered minerals, such as clays, carbonates and sulphates, yielded nothing definitive. From the geological evidence, and from climate models that consistently implied enduring and intensely cold conditions, it was difficult to escape the conclusion that Mars had almost always been in a deep freeze.

Then, in the late 1990s, the Thermal Emission Spectrometer on board the orbiting Mars Global Surveyor detected small patches of grey haematite in isolated locations on the surface³. This kind of haematite almost always requires liquid water for its formation. Its signature was the siren song that lured the Opportunity rover to Mars' Meridiani Planum (Fig. 1), a flat, volcanic and sedimentary plain on Mars just east of the giant canyon system Valles Marineris. Opportunity is one of two Mars Exploration Rovers whose mission has been to look for geological and geochemical signs that Mars may once have had an environment conducive to life. What was missing until the rovers' expedition was definitive

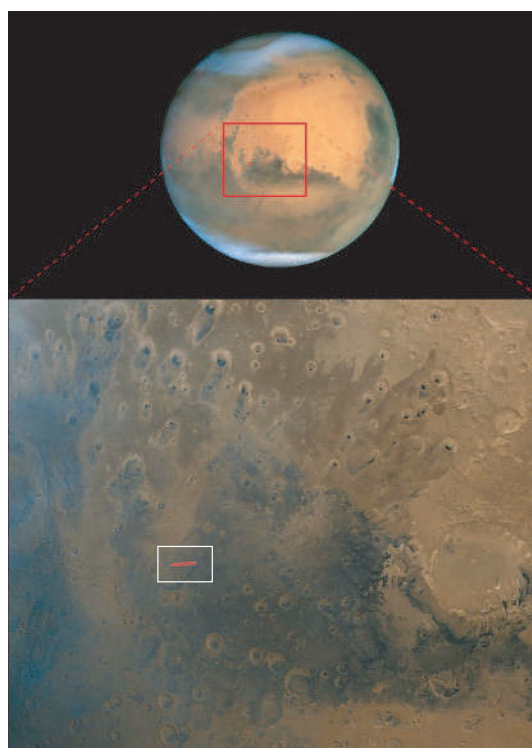


Figure 1 | Opportunity's site of operation. Top, an image of Mars taken by the Hubble Space Telescope during Mars' closest approach to Earth in June 2001. Bottom, the Meridiani Planum region compiled from Viking Orbiter images taken in 1980 during mid-northern summer on Mars. The red ellipse on the left is about 87 km long and marks the landing zone of the Opportunity lander. The large circular feature on the right is the Schiaparelli impact crater.

evidence that water had at one time had a significant chemical role in the evolution of the martian surface, as it has had on Earth.

What Opportunity has achieved, from the moment it swung into action on 25 January 2004, has been stunning. It landed on Mars within metres of a rock outcrop, and detailed analyses of this and other outcrops at many locations over its 2-kilometre (and counting) traverse across the plains of Meridiani show that they are composed of layered sandstones. Those sandstones are primarily made of mixtures of magnesium sulphates, iron sulphates and silicon-rich, sand-sized particles of rock.

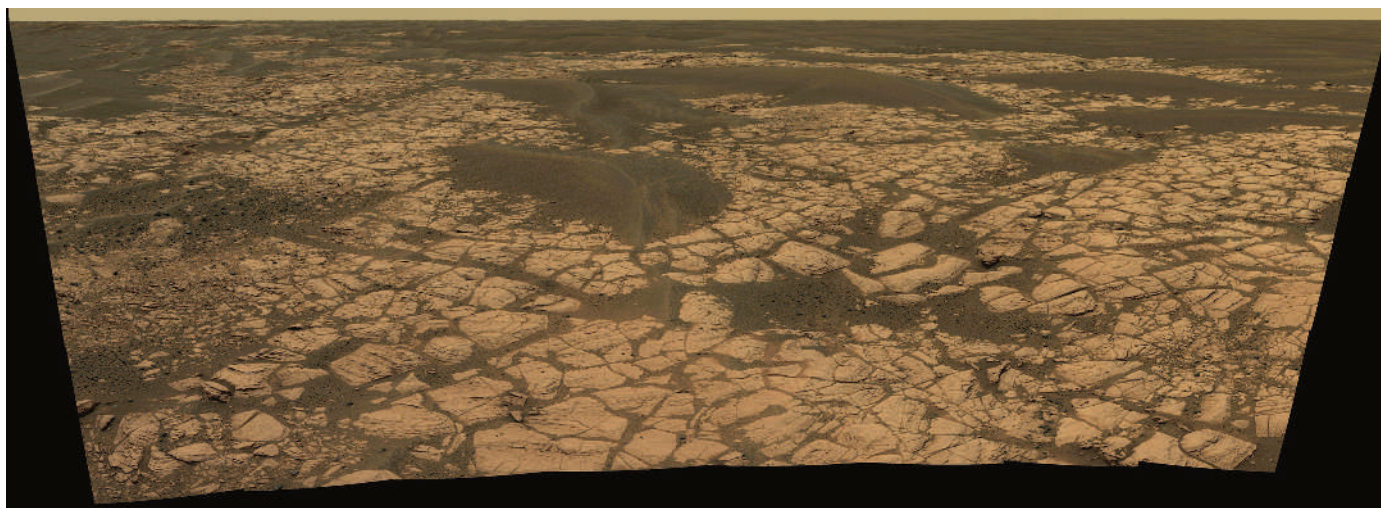
Dispersed across the plains and within the sandstone matrix are enigmatic nodules of almost pure haematite⁴.

The rovers' team has been consumed with trying to understand how the sulphate-rich layers were formed. The most comprehensive data so far have come from a 7-metre-high exposure within Endurance crater, some 800 metres from Opportunity's landing site. Here, the team has documented a complex history of events from a careful analysis of the stratigraphy, textures and composition of the entire exposure. Remarkably, these investigations at Endurance were completed after about 320 sols, or martian days, four times longer than the design lifetime of the rovers. The team concludes that the sandstones were formed by the erosion and redeposition of fine-grained silicate particles and evaporites that were derived from the chemical weathering of volcanic rocks by acidic waters. These volcanic rocks, called olivine basalts, are iron- and magnesium-rich silicates that are known to crystallize first from a molten magma source. They are commonly found at terrestrial hot-spot volcanic sites on Earth, such as Hawaii.

The uppermost sections of the Endurance exposure exhibit cross-bedding, which indicates deposition in shallow waters that once existed in a playa-like setting between sand dunes. Jarosite, an iron sulphate that forms only at extremely low pH, probably precipitated from these waters, and the variety of intergranular cements, haematite concretions and crystal 'moulds' attests to multiple episodes of inundation resulting from changes in groundwater levels⁵.

This is the conclusion challenged by the papers in this issue^{1,2}. McCollom and Hynke¹ (page 1129) hypothesize that the deposits seen at Meridiani were instead produced by the deposition of volcanic ash, followed by alteration of that ash by small amounts of acidic water and sulphur dioxide. Their primary observation, from an analysis of the data from Eagle crater during the first 45 sols of the

NASA/HUBBLE HERITAGE TEAM/US GEOL. SURV.



NASA/JPL/CALTECH/CORNELL UNIV.

Figure 2 | Sulphate, sulphate everywhere. Opportunity's view of a vast field of sulphate-rich sedimentary rocks encountered on the way from Endurance crater to Victoria crater. This is an outcrop dubbed Olympia, near the Erebus crater. The view is to the south, in the direction of Opportunity's traverse. Outcrops such as these are common along the traverse, attesting to the large extent of these deposits.

mission, is that the composition of the outcrops seen at Meridiani seems very like that of typical martian basalts (measured both *in situ* and in the Shergotty meteorites found on Earth) with sulphur added. They point out that any model of the chemistry of the Meridiani rocks must explain why the rocks are enriched in sulphate but not in any major cations — if the sulphate were attributable to precipitation of salts from an evaporating brine, the rocks would be enriched in a balancing cation such as calcium, magnesium or iron. This is not observed. However, this reasoning is valid only if the silicate particles in the outcrop are themselves unaltered. Cations could have been removed from the silicate portion by acid weathering before incorporation in the outcrop, and this would not have been detectable by the rover instruments. McCollom and Hynek's explanation for the exposed bedrock is that it was originally a basaltic ash deposit resulting from an explosive volcanic process, and was subsequently altered through reaction with an aqueous sulphuric acid solution derived from condensation of vapours rich in sulphur dioxide and water.

The patterns in the features seen at the Meridiani outcrops are also observed in volcanic ash deposited by surges of explosive volcanism on Earth. If McCollom and Hynek's scenario for the formation of the Meridiani deposits is correct, the origin and modification of these sediments would have occurred at high temperature with little groundwater (and no surface water), greatly reducing the possibility that these rocks indicate that a habitable environment ever existed at Meridiani.

In the second paper in this issue, Knauth *et al.*² (page 1123) propose a scenario for the origin of the Meridiani sulphate deposits that is similarly pessimistic about the evidence for past water. Their explanation is that the deposits were produced by a ground-hugging, turbulent surge of rock fragments, salts, sulphides, brines and ice produced by the impact

of a meteorite. Subsequent weathering by intergranular water films could then account for all of the features observed, without invoking the existence of shallow seas, lakes or near-surface aquifers. It is possible that the layers traversed by Opportunity resulted from one impact event, possibly the one that produced the 450-km-wide Schiaparelli crater lying about two crater diameters to the east (Fig. 1).

Knauth *et al.* note that the patterns of sedimentary structures created by surges closely resemble those produced by wind and deposition in a shallow body of water. Because of the complexity of the flow of ejecta generated by an impact, a remarkably wide range of depositional conditions can develop that look very much like those found in other sedimentary environments. Once in place, the heterogeneous jumble of phases would undergo alteration by small amounts of interstitial waters.

All interpretations of scientific results from other worlds have the same difficulty: extrapolations from terrestrial experience must be made from limited spacecraft data. Earth's geological history is rich and complex, and that of Mars must surely have been as well. The two papers in this issue have much in common, especially given the ultimate goal of the Mars rovers — to search for evidence of past habitable conditions on Mars. McCollom and Hynek¹ suggest volcanic-ash deposits as the source of the sulphate-rich outcrops, and Knauth *et al.*² propose a similarly dry phenomenon — that turbulent ejecta from meteorite impacts could have been the culprit. Both groups propose scenarios that preclude the existence of significant bodies of water at the surface (at least at Meridiani), and therefore that Mars may never have had conditions conducive to life. This conclusion stands in sharp contrast to the provocative interpretation that there must have been long-lived surface water to form the Meridiani outcrops⁵. Given how enticing this latter interpretation is, it is vital to explore alternative possibilities

whatever the ultimate verdict proves to be.

It is clear that the sulphate-rich rocks seen along Opportunity's path are not simply a fortuitous, local discovery. Such rocks, in the form of massive light-toned stacks about a kilometre thick, seem to underlie most of the hundreds of thousands of square kilometres of the Meridiani region⁶. Sulphates have also been seen in extensive dune fields in the north polar regions, and within the vast canyon system of Valles Marineris, by the OMEGA instrument on board the Mars Express spacecraft. Any reconstruction of the history of water on Mars must explain the existence of these massive sedimentary deposits, and the observations by OMEGA that clays are also abundant on Mars⁷.

Perhaps the most wondrous aspect of this bold new era of Mars exploration is the robustness and versatility of our robotic explorers. With perseverance and luck, we can hope for the same detailed stratigraphic analysis performed at Endurance crater⁸ to be carried out at more distant locales. In particular, Opportunity is making steady progress to the southern reaches of Meridiani (Fig. 2), and will hopefully reach the much larger Victoria crater. Whatever new vistas are in prospect, they are sure to provide fresh data and ideas on the history of water on Mars, and new fodder for scientific debate on the planet's habitability. ■

Mark A. Bullock is at the Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, Colorado 80302, USA.

e-mail: bullock@boulder.swri.edu

1. McCollom, T. M. & Hynek, B. M. *Nature* **438**, 1129–1131 (2005).
2. Knauth, L. P., Burt, D. M. & Wohletz, K. H. *Nature* **438**, 1123–1128 (2005).
3. Christensen, P. R. *et al.* *J. Geophys. Res.* **105**, 9623–9642 (2000).
4. Squyres, S. W. *et al.* *Science* **306**, 1709–1714 (2004).
5. Squyres, S. W. *Earth Planet. Sci. Lett.* **240**, 1–10 (2005).
6. Hynek, B. M., Arvidson, R. E. & Phillips, R. J. *J. Geophys. Res.* **107**, 5088; doi:10.1029/2002JE001891 (2002).
7. Poulet, F. *et al.* *Nature* **438**, 623–627 (2005).
8. Grotzinger, J. P. *Earth Planet. Sci. Lett.* **240**, 11–72 (2005).

PHYSICS

Philately will get you everywhere

The World Year of Physics 2005 did not fail to make its mark — its postmark, at least. This page displays a collection of the stamps and associated philatelic memorabilia issued by the world's postal services to commemorate the occasion (see also <http://fizjlk.fic.uni.lodz.pl/rut/Stamps/wyp/wyp2005.html>).

The event may have been global, but the iconography of the stamps is, in many cases, distinctly national. The Republic of Ireland, for example, takes the opportunity to celebrate the 200th birthday of William Rowan Hamilton, the prodigious mathematician and physicist. Apocryphally, the formula for the multiplication of quaternions (four-dimensional complex numbers used in an early form of vector algebra) depicted on the Irish 48-cent stamp came to Hamilton as he was walking along the Royal Canal in Dublin. He carved it into the stone of the nearby Broome Bridge — an intellectual cut above most graffiti.

Slovakia likewise commemorates one of its own, Dionýz Ilkovič, with a first-day cover that includes his expression for the mean-limiting diffusion current in polarography, an electrochemical analysis technique. India trumps this with a triumvirate: Satyendra Nath Bose, of boson fame; Homi J. Bhabha, the nuclear physicist who lent his name to electron-positron scattering; and Subrahmanyan Chandrasekhar, the astrophysicist who looked into black holes.

Unsurprisingly, however, the iconic status of one physicist transcends national boundaries. The World Year of Physics does, after all, celebrate the hundredth anniversary of Albert Einstein's *annus mirabilis*, during which he published five papers — covering atomic behaviour, the quantization of light, and the nature of space and time — from which classical physics

never recovered. The most famous product of that fertile year, the equation that embodies mass-energy equivalence, is no less iconic than its creator — though not, admittedly, as $m = E/c^2$, the form in which it appeared in 1905.

Many countries are eager to stamp their claim on the greatest physicist of the twentieth century: Germany, naturally, where he was born in 1879 and whence he fled in 1933; and Switzerland, scene of his greatest triumphs, which famously emanated from the patent office at Bern. The Czech Republic remembers his association with Prague, where he obtained his first full professorship in 1911, and where he consorted with the writers Max Brod and Franz Kafka in Bertha Fanta's salon on the Altstädter Ring. And Italy chooses for its 85-cent stamp (next to a Feynman diagram and a depiction of the birth of a black hole) a woodcut of the university town of Pavia, near Milan, where Einstein's family settled in his late teens. Cuba even breaks with the Year of Physics theme, commemorating instead the seventy-fifth anniversary of a visit by Einstein to the island.

With his luxuriant moustache and untidy hair, and expression ranging from the lugubrious to the avuncular, Einstein — preferentially depicted as an old man here — is instantly recognizable. This remains so even when pictured with half his head missing (Poland), or when the depiction is schematic in the extreme (Cuba, Israel). Which raises the question, is there a minimal Einstein? What are the bare essentials required to make an image that remains unmistakably him? Readers might care to send in their own drawings, with contact details, by fax to +44 (0) 20 7843 4596 or as a pdf attachment to minimaleinstein@nature.com

Richard Webb



MOLECULAR BIOLOGY

Antagonizing the neighbours

Joel C. Eissenberg and Sarah C. R. Elgin

Nucleosomes bundle up the DNA in a cell's nucleus, wrapping it around a complex of histone proteins. Studies of histone modifications and the proteins that bind to them reveal a mechanism that may control this packing.

Crack open any cell nucleus and look inside: you'll see what look like beads on a string. The beads are nucleosomes, small protein complexes that help to package the DNA (the strings) into the cramped confines of the nucleus. In the past fifteen years, nucleosomes have graduated in our understanding from being passive spools for DNA to full partners in the control of genetic information in the cell. Diverse chemical modifications of the histone proteins that form the nucleosome core can alter the expression of the associated genes¹. These modifications make up what is known as the histone code, and a major challenge in molecular biology is to decipher how they affect gene expression.

Two of the most common modifications are phosphorylation and methylation — respectively the addition of a phosphate or a methyl group to the amino acids of which the histones are composed. Allis and colleagues² previously proposed that reversible phosphorylation of the amino acids serine or threonine in the 'tail' regions of histones could antagonize the binding of regulatory proteins to neighbouring methylated lysine amino acids, creating a binary control switch. In this issue, Fischle *et al.* (page 1116)³ and Hirota *et al.* (page 1176)⁴ present data that strongly support this model and advance our understanding of how histone modification can control chromosome function.

Among the many histone modifications that have been described, methylation of lysines and phosphorylation of serines and threonines have attracted much attention. Phosphorylation of the serine at the tenth position in the tails of histone H3 (H3S10)

occurs during cell division in eukaryotic (higher) cells. Once the cells have replicated their DNA and begin to prepare for division, nearly all of the histone H3 in the nucleus seems to be phosphorylated at this site. Phosphorylation of H3S10 also occurs at other stages in the cell cycle, but only at discrete chromosomal sites that are associated with gene expression.

The methylation of lysines is more complex. Methylation at lysine 9 of histone H3 (H3K9) is found mainly in the heterochromatin — the dense, mostly inactive regions of the genome. Methylation at lysine 4 of histone H3 (H3K4), by contrast, is associated with active genes. The different outcomes of lysine methylation result from the fact that each modification creates a binding target for a distinct protein. Heterochromatin protein 1 (HP1), which promotes heterochromatin formation (and the consequent gene silencing), recognizes and binds to methylated H3K9 using a region called the chromodomain^{5,6}. CHD1, an enzyme that may destabilize nucleosomes and expose the DNA for gene expression, recognizes and binds to methylated H3K4 through two tandem chromodomains⁷.

Fischle *et al.*³ and Hirota *et al.*⁴ examined cells progressing into the metaphase stage of cell division, where the chromosomes become very tightly packed, or 'condensed', to facilitate their separation into the future daughter cells. Both groups discovered that these cells accumulated histone H3 that is both methylated at lysine 9 and phosphorylated at the neighbouring serine 10. Using antibodies specific for the doubly modified H3, the authors found that this dual modification occurred specifically in

heterochromatic regions of chromosomes.

During most of the cell cycle, HP1 is also concentrated in heterochromatin, but with the onset of chromosome condensation, much of the HP1 leaves the chromosomes. Both labs link the dissociation of HP1 with the accumulation of phosphorylated H3S10 by showing that inactivation of an enzyme that phosphorylates H3S10 (Aurora B kinase) results in the retention of HP1. These observations suggest that H3S10 phosphorylation causes displacement of HP1 from heterochromatin with the onset of metaphase. To test this, the ability of HP1 to bind to dually modified H3 tail peptides was measured *in vitro*, using either fluorescence polarization³ or binding to peptide-coated beads⁴. In both assays, the binding of HP1 to the methylated H3K9 was substantially impaired when the neighbouring H3S10 was simultaneously phosphorylated.

The results demonstrate the need to examine the impact of multiple modifications on a histone tail; using an antibody that recognizes a single modification is clearly not sufficient to infer the histone state. The findings also provide experimental support for the regulatory-switch hypothesis of Allis and colleagues². Methylation at H3K9, and concomitant binding of HP1, is not only found in heterochromatin, but also contributes to inactivation of some genes in euchromatin (the less dense and more active regions of the genome)⁸. If HP1 can be evicted by phosphorylation of H3S10, such repression might be reversed. However, the genomic region where HP1 binds would still be tagged by the methyl groups at H3K9. So if the H3S10 phosphate group were removed (by a protein phosphatase), that would leave the unopposed H3K9 methyl mark available to restore HP1 binding and reconstitute heterochromatin. This model is consistent with genetic studies implicating a histone methylase^{9,10} and a protein phosphatase¹¹ in the control of heterochromatin formation in the fruitfly. Whether it applies to regulation in euchromatic domains remains to be seen.

Another study reported in this issue (page 1181)⁷ suggests that the methyl-phospho

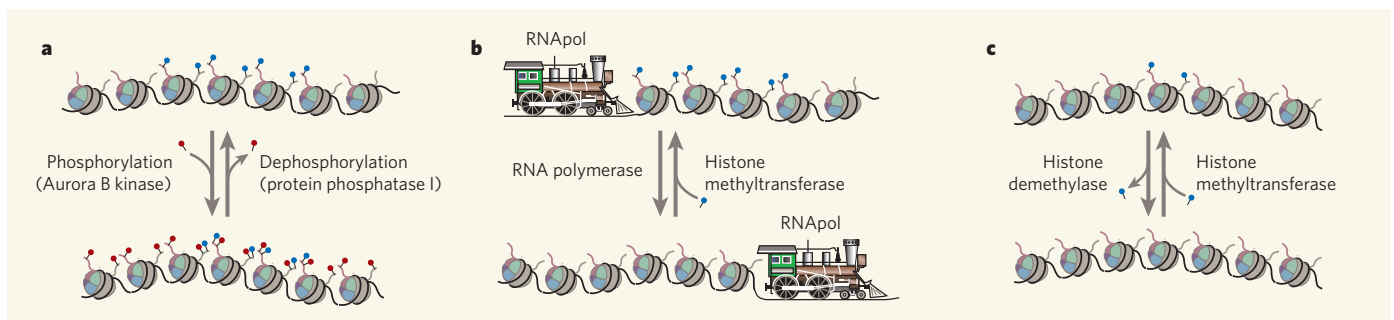


Figure 1 | Three-pronged control. Three contrasting mechanisms to regulate binding of proteins that target methylated histones. **a**, Fischle *et al.*³ and Hirota *et al.*⁴ show that phosphorylation of a serine or threonine amino acid that lies next to a methylated lysine creates a methyl-phospho module. This module can no longer associate with methyl-binding proteins. Phosphorylation is readily reversed by a phosphatase. This mechanism

leaves the original pattern of methylation intact. **b**, Passage of RNA polymerase II may result in nucleosome replacement, erasing the original pattern of methylation. **c**, Site-specific patterns of methylated nucleosomes can be erased enzymatically by histone demethylase. In cases **b** and **c**, the pattern of methylation can only be restored by targeted *de novo* histone methyltransferase activity.

switch mechanism has wider applicability. Flanagan *et al.* describe the crystal structure of the double chromodomains of the mammalian CHD1 protein bound to an H3 peptide containing methylated H3K4. The way in which the CHD1 chromodomains bind to methyl-lysine seems different from how HP1 binds. But binding of the CHD1 chromodomains to methylated H3K4 is antagonized *in vitro* by phosphorylation of a neighbouring threonine (H3T3). CHD1 resembles a helicase, an enzyme capable of loosening nucleosome–DNA contacts. So controlled interaction of CHD1 with H3K4 by phosphorylation of H3T3 might regulate the access of regulatory proteins to DNA to control gene expression.

Using a binary switch to eject proteins that bind to methylated histones essentially reverses the effects of histone methylation. So far, the cell has been found to use two other methods to accomplish this end. First, as RNA polymerase traverses genes to make the encoded messenger RNA, it displaces the nucleosomes from the DNA. Nucleosomes re-form once the polymerase has passed, and the process can replace methylated histones with unmethylated ones¹². Second, histone demethylase enzymes can directly strip the methyl groups off specific lysines^{13,14}.

Why use three mechanisms to achieve the same biochemical result? Each has a different overall impact. Eviction of nucleosomes by RNA polymerase can remove all nucleosome modification marks (Fig. 1b). Lysine demethylation can target specific nucleosomes, but will also erase methylation patterns (Fig. 1c). However, by leaving methyl marks intact and ejecting the methyl-lysine binding factors through phosphorylation of neighbouring amino acids, cells can rapidly effect a large-scale reorganization of chromosome structure while preserving the underlying methylation pattern (Fig. 1a); this allows the original structure to be restored on dephosphorylation.

Fischle *et al.*² identified 16 instances of lysines flanked by either serine or threonine among the four histone proteins that form the nucleosome, suggesting the possibility of other such binary switches. The monotonous beads-on-a-string conceal a rich variety of ornaments that compete with one another to control gene expression. ■

Joel C. Eissenberg is in the Edward A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, St Louis, Missouri 63104, USA. Sarah C. R. Elgin is in the Department of Biology, Washington University, St Louis, Missouri 63130, USA.
e-mails: eissenjc@slu.edu;
selgin@biology.wustl.edu

5. Jacobs, S. A. & Khorisanizadeh, S. *Science* **295**, 2080–2083 (2002).
6. Nielsen, P. R. *et al.* *Nature* **416**, 103–107 (2002).
7. Flanagan, J. F. *et al.* *Nature* **438**, 1181–1185 (2005).
8. Nielsen, S. *et al.* *Nature* **412**, 561–565 (2001).
9. Tschiersch, B. *et al.* *EMBO J.* **13**, 3822–3831 (1994).

10. Schotta, G. *et al.* *EMBO J.* **21**, 1121–1131 (2002).
11. Baksa, K. *et al.* *Genetics* **135**, 117–125 (1993).
12. McKittrick, E., Gafken, P. R., Ahmad, K. & Henikoff, S. *Proc. Natl Acad. Sci. USA* **101**, 1525–1530 (2004).
13. Shi, Y. J. *et al.* *Cell* **119**, 941–953 (2004).
14. Metzger, E. *et al.* *Nature* **437**, 436–439 (2005).

ATMOSPHERIC PHYSICS

Reflections on aerosol cooling

Jim Coakley

By changing the composition of Earth's atmosphere, human activity has both a warming and a cooling effect on the planet. According to new calculations, that latter influence is large, but it is likely to be declining.

The Earth is warming because of rising concentrations of greenhouse gases in the atmosphere. But some of the human activities that produce those gases, including burning of fossil fuels and biomass, also produce hazes that partially offset the warming. The extent of this cooling influence is not known. But it occurs because human-generated haze particles scatter and absorb incoming sunlight, an effect known as 'aerosol direct radiative forcing' that reflects solar radiation back into space.

On page 1138 of this issue, Nicolas Bellouin and colleagues¹ provide a new estimate of this aerosol forcing effect. The figure they produce ($0.8 \pm 0.1 \text{ W m}^{-2}$ of incident sunlight) is at the high end of the range ($0.2\text{--}1.0 \text{ W m}^{-2}$) given in the previous report from the Intergovernmental Panel on Climate Change (IPCC)². Notably, this new estimate comes with an uncertainty of only 10–15%, a far cry from the range cited in the IPCC assessment. For comparison, the warming effect from the build-up of the long-lived greenhouse gases carbon dioxide, methane, nitrous oxide and chlorofluorocarbons is $2.4 \pm 0.2 \text{ W m}^{-2}$.

Because of concerns about human health and acid rain, considerable efforts have been and are being made to improve air quality and reduce the amounts of anthropogenic aerosols that arise, for example, from power-plant and manufacturing emissions. So an implication of the figure obtained by Bellouin *et al.* is that such improvements will come at the price of substantial additional warming — as much as half that already experienced from the build-up of greenhouse gases.

Estimating the effect of anthropogenic aerosols on climate is notoriously difficult. Computer models must predict the concentrations, and the physical and chemical make-up of the particles, and the consequent radiative properties. The variability of estimates for the aerosol direct radiative forcing in the IPCC report reflects the wide range of possible outcomes. To constrain this range, Bellouin *et al.* took advantage of new observations of aerosol properties from the Moderate Resolution Imaging Spectrometer (MODIS) instrument

now flying on NASA's Terra and Aqua satellites.

A central concept in these studies is 'aerosol optical depth', a measure of the attenuation of sunlight by particles that is proportional to the amount of aerosol. Over oceans, the MODIS observations separate the aerosol optical depths into the fractions contributed by small and large particles³. *In situ* measurements and surface-based observations of light attenuation and scattering led Bellouin *et al.* to propose using the separation of the optical depth into contributions made by small, anthropogenic particles and large, natural particles — such as windblown dust and sea spray — to estimate what fraction of aerosols is attributable to anthropogenic causes.

Anthropogenic aerosols over the oceans account for only one-third of the aerosol direct radiative forcing, however. Haze over continents accounts for the rest. But MODIS observations are more uncertain over land, and the separation of aerosol optical depth into fractions contributed by small and large particles is not feasible³. As an alternative, Bellouin *et al.* divide the continents into six parts and use predictions from an ensemble of aerosol chemical-transport models to determine the fraction of the total aerosol optical depth contributed by anthropogenic aerosols. They use optical properties for the particles derived from surface-based observations of the scattering and attenuation of sunlight⁴, the data coming from one observing site for each of the six continental regions. The particle properties from the surface observations are combined with the aerosol optical depth from the MODIS observations, and the fraction attributed to humans from the models, to predict — through radiative transfer calculations — the effect of anthropogenic aerosols on reflected sunlight.

To estimate the uncertainty in their calculations, Bellouin *et al.* performed a large number of Monte Carlo simulations. In these simulations, the various parameters that affect the aerosol direct radiative forcing — such as the MODIS aerosol optical depth, the model predictions of anthropogenic fractions and the surface-based estimates of the aerosol optical

1. Khorisanizadeh, S. *Cell* **116**, 259–272 (2004).
2. Fischle, W., Wang, Y. & Allis, C. D. *Nature* **425**, 475–479 (2003).
3. Fischle, W. *et al.* *Nature* **438**, 1116–1122 (2005).
4. Hirota, T., Lipp, J. J., Toh, B.-H. & Peters, J.-M. *Nature* **438**, 1176–1180 (2005).



50 YEARS AGO

"Influence of space flight on engineering and science"

— Within the past few years many scientists have predicted seriously and confidently that human beings from the Earth would, in the foreseeable future, travel to the Moon and the nearer planets. The ranks of those who would dispute this project are diminishing rapidly. Although much of the progress is still guarded by military necessity, space flight is emerging as an activity in its own right — one that can command the efforts of many scientists and engineers... A recent survey shows that the study of physics in American public high schools has been declining for more than half a century... why [does our youth] turn away from a career in science? We can only grope for the answer. Perhaps they sense, better than their elders, that too much of our scientific talent is engaged in the unproductive task of developing weapons for war. Is there much inspiration to devote one's life to this end, when we are rapidly approaching the borderline of total destruction? I believe that space flight might serve in no small measure to turn men's minds toward a more appealing scientific goal. As the exploits of Cabot, Drake and Davis inspired many generations of Englishmen to turn to the sea, so may the first astronauts reawaken our youth to the romance of scientific exploration.

Milton W. Rosen, Naval Research Laboratory, Washington, D.C.
From *Nature* 24 December 1955.

100 YEARS AGO

Heredity. By C. W. Saleeby, M. D. The appearance of a little shilling book on heredity is almost startling, when we consider the difficulty of the subject and the relative youth of its exact study. That a book like this should be possible indicates that considerable progress has been made in recent years. Was it not Leibnitz who said, "The more a science advances, the more it becomes concentrated in little books"?

From *Nature* 21 December 1905.

properties — were 'stepped' through their ranges of uncertainties. This process produced probability distribution functions that represent the probable range of the forcing. The relatively small uncertainty reported by Bellouin *et al.* arises from use of the relatively accurate MODIS optical depths, as compared with the wide range of optical depths generated by the aerosol chemical-transport models that contributed to the IPCC assessment.

So far, so good. But this won't be the end of the story. For example, one wonders how well global estimates of biases in the MODIS aerosol optical depths³, which Bellouin *et al.* attempted to remove, coupled with the aerosol optical properties derived from just six continental sites, characterize aerosols of anthropogenic rather than natural origin. Also, Bellouin *et al.* assumed that the aerosol direct radiative forcing for overcast regions was negligible. As they note, such forcing will be difficult to deduce, but it is bound to be as large as, if not greater than, their claimed uncertainty.

Likewise, the MODIS aerosol optical depth increases with increasing cloud cover⁵, whereas the comparisons with surface-based observations used to establish the accuracy of the MODIS aerosol properties favour largely cloud-free conditions³ — changes in aerosol properties in the vicinity of clouds suggest that the MODIS observations could have biases that have not yet been characterized. Finally, aerosols also affect the size and numbers of droplets in clouds, thereby altering the amount

of sunlight reflected by clouds. The extent of this effect, known as 'aerosol indirect radiative forcing', remains largely unknown. But it may offset greenhouse-gas warming even more than the aerosol direct radiative forcing⁶.

Assessments of climate change caused by human activity have been stymied in part by the sizeable uncertainty in estimates of the aerosol direct and indirect radiative forcings. The strategy of using combinations of global space-based and surface-based observations to constrain model estimates, as followed by Bellouin *et al.*, is a promising way of reducing these uncertainties. Space missions such as CALIPSO and CloudSat are to become part of the A-Train — the Aqua satellite constellation — early next year. They will help to improve the characterization of aerosols, particularly over continents where the direct radiative forcing is greatest, as well as the treatment of cloud-aerosol interactions.

Jim Coakley is at the College of Oceanic and Atmospheric Sciences, 104 COAS Admin Building, Oregon State University, Corvallis, Oregon 97331-5503, USA.
e-mail: coakley@coas.oregonstate.edu

1. Bellouin, N., Boucher, O., Haywood, J. & Reddy, M. S. *Nature* **438**, 1138–1141 (2005).
2. Ramaswamy, V. *et al.* (eds) *Climate Change 2001: The Scientific Basis. Contribution of WG1 to the Third Assessment Report of the IPCC* (Cambridge Univ. Press, 2001).
3. Remer, L. A. *et al.* *J. Atmos. Sci.* **62**, 947–973 (2005).
4. Dubovik, O. *et al.* *J. Atmos. Sci.* **59**, 590–608 (2002).
5. Loeb, N. G. & Manalo-Smith, N. J. *Clim.* **18**, 3506–3526 (2005).
6. Anderson, T. *et al.* *Science* **300**, 1103–1104 (2003).

GENOMICS

Multiple moulds

André Goffeau

Three species of *Aspergillus* fungi are the latest organisms to have their genome sequenced. Comparison of the genomes sheds light on, among other things, what endows them with pathogenic or beneficial features.

The genome sequences of three *Aspergillus* fungi are reported in this issue: *Aspergillus oryzae*¹, used in making the Japanese drink sake; the human pathogen *Aspergillus fumigatus*²; and the genetic model species *Aspergillus nidulans*³. The 185 known species of *Aspergillus* include 20 human pathogens, numerous plant pathogens and a variety of species that we use to produce foods, chemicals and industrial enzymes. The genomes provide a wealth of information about the evolution of this fascinating group of organisms, and about the beneficial or detrimental characteristics of each species.

The sequences, published by teams from Japan, the United States and Europe, cover an average of nearly 95% of each genome. In total across the three species, more than 95 megabases have been sequenced, crammed with

over 33,500 protein-coding genes contained on 24 chromosomes (eight chromosomes per species). By comparison, the human genome has about 30,000 protein-coding genes in 3,000 megabases.

Aspergillus oryzae has been used for nearly a thousand years to produce traditional Japanese fermented foods and drinks. Its genome¹ has about seven to nine megabases more DNA than *A. fumigatus* and *A. nidulans*. To account for this, the authors propose that some genes were transferred to *A. oryzae* from other species during evolution. The extra DNA stretches are dispersed throughout the genome and are enriched in genes involved in the synthesis and the transport of numerous secondary metabolites — the chemical compounds in an organism that are not directly involved in normal growth, development or reproduction.

Secondary metabolites are often specific to one or a few species, so they provide a window on the particular biology of the species.

Species closely related to *A. oryzae*, such as *Aspergillus flavus* and *Aspergillus niger*, have similar gene acquisitions. For instance, the toxic *A. flavus* has 25 genes encoding proteins involved in the pathway that produces the poisonous 'aflatoxin'. These genes are present in *A. oryzae* but are not expressed. It is likely that an ancestor of *A. flavus* passed these genes to *A. oryzae*, and that they were then inactivated during the subsequent evolution of *A. oryzae*.

Aspergillus fumigatus is a potentially deadly human pathogen and a major allergen. The *A. fumigatus* sequence² pinpoints nine previously unknown allergens, numerous genes involved in the production of specific secondary metabolites, and a set of essential genes that may be potential targets for drugs. However, the factors that underlie the pathogenicity of this species are complex, and their identification required other approaches to complement the genome analysis. For instance, for *A. fumigatus* to thrive inside warm-blooded creatures such as ourselves, it must be able to tolerate our high body temperature (compared with that of the external environment). Using DNA microarray analysis, a set of 'thermotolerance genes' whose activity increases at 37 °C has been identified. But it seems that warming up to 37 °C is insufficient to turn on many genes that are associated with virulence in this species.

Aspergillus nidulans has long been a model organism used to study the genetics of fungi. Its genome³ was crucial for the comparative analysis of the three aspergilli, but it also had some features of its own to reveal. For instance, the regulation of several of its genes was clarified, with the identification of putative binding sites for gene regulatory factors and control elements, as well as many short open reading frames that lie upstream of genes; these short sequences may stall the expression of neighbouring genes. In addition, the sequence disclosed many previously unknown genes involved in peculiar metabolic (fatty-acid oxidation), developmental (polarized growth) and DNA-repair pathways.

The three species diverged several hundred million years ago, and their genomes differ considerably³. There are almost 3,000 proteins that are closely related, or 'homologous', among the genomes. On average, these proteins have only 68% of their constituent amino acids present in all three genomes — a value comparable to that of proteins homologous between mammals and fish, which diverged around 450 million years ago. Nevertheless, the order of the homologous proteins along chromosomes (synteny) is conserved in the three species, indicating that no whole-genome duplication occurred during evolution. However, large regions lack any synteny because of small tandem repeats, gene rearrangements in the chromosome extremities, and considerable random breakage and

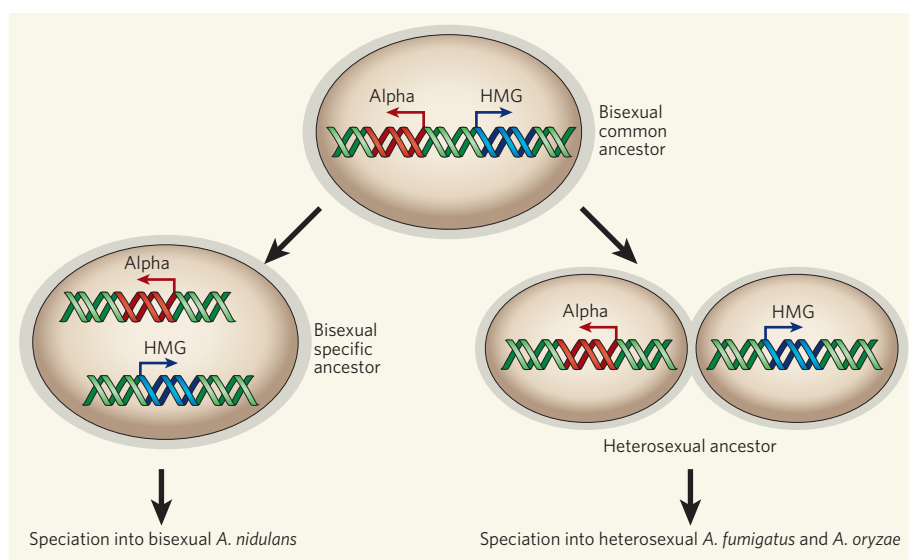


Figure 1 | Evolutionary model of the *Aspergillus* mating genes. In a bisexual common ancestor, the two mating-type genes, alpha (red) and HMG (blue), were fused head to tail on the same chromosome and share a similar flanking, regulatory region (green). In the bisexual *A. nidulans*, the chromosome is broken and the two mating-type genes end up on different chromosomes (with their flanking regions). In the ancestor of *A. oryzae* and *A. fumigatus*, the two mating-type genes dissociate in different strains, but remain flanked by similar genes. After speciation, both *A. fumigatus* and *A. oryzae* become fully heterosexual, with some isolates having only the alpha mating type and others only the HMG mating type, both being in similar chromosomal environments.

rearrangements of syntenic blocks. Such genome reorganization is seen to a greater extent in *A. oryzae* than in *A. fumigatus*. The rates of amino-acid evolution within homologous genes are similar for all three species, so the evolution of large structural rearrangements does not parallel the rate of individual amino-acid changes.

The chief revelation from the three-genome comparison is the mating systems in *A. fumigatus* and *A. oryzae*. Sexual reproduction in yeast can take place only between individuals of opposite mating type, or 'sex', as determined by mating-type genes. *Aspergillus nidulans* has two mating-type genes: one contains an alpha box and the other a high-mobility-group (HMG) domain. So each cell can have two sexes at once, and *A. nidulans* is self-fertile. *Aspergillus nidulans* can also reproduce asexually by 'mitotic reproduction', creating spores that are sprinkled by structures known as conidiophores.

Aspergillus fumigatus and *A. oryzae* were believed to reproduce only through the asexual mitotic process. Unexpectedly, however, the *A. oryzae* and *A. fumigatus* genomes each have a mating-type gene: the *A. oryzae* sequence contains an alpha mating-type gene, whereas the *A. fumigatus* sequence has an HMG mating-type gene. These genes occupy nearly identical positions in their respective genomes, with conserved synteny for 1.7 megabases on either side. In addition, 215 genes implicated in different phases of the *A. nidulans* mating process occur in *A. oryzae* and *A. fumigatus*. These and other recent data⁴ raise the possibility that *A. fumigatus* and *A. oryzae* are heterosexual, and that conversion of bisexuality to heterosexuality occurred during

the evolution of the *Aspergillus* genus (Fig. 1).

These reports describe only the initial examination of the genomes, of course, and the sequences provide much scope for further analyses. The sequencing of other *Aspergillus* genomes is under way and will provide an even broader perspective on the biology and evolution of these fungi. The most keenly anticipated *Aspergillus* sequence is that of *A. niger*, which has long been used in the industrial production of citric acid⁵. The commercial significance of several *Aspergillus* species has meant that their genome sequences, including that of *A. niger*, have been kept behind the closed doors of biotechnology companies for some time. However, this practice seems to be changing: a consortium of Japanese companies has agreed to release its *A. oryzae* sequence, and Monsanto provided access to its *A. nidulans* genome sequence, so that they could be added to the publicly funded sequences now published. And, fortunately, the US Department of Energy has undertaken to complete one of the industrial *A. niger* sequences (which is currently of low coverage) to make public a useful version of this genome. Perhaps the time when genome sequences belong exclusively to industry is over.

André Goffeau is at the Institut des Sciences de la Vie, Université catholique de Louvain, Louvain-la-Neuve 1348, Belgium.
e-mail: goffeau@fyssa.ucl.ac.be

1. Nierman, W. C. et al. *Nature* **438**, 1151–1156 (2005).
2. Machida, M. et al. *Nature* **438**, 1157–1161 (2005).
3. Galagan, J. E. et al. *Nature* **438**, 1105–1115 (2005).
4. Paoletti, M. et al. *Curr. Biol.* **15**, 1242–1248 (2005).
5. Hennebert, G. L. *Recueil de Travaux d'Histoire et de Philologie* **6**, 61–103 (Univ. Louvain, 1979).

OBITUARY

Richard E. Smalley (1943–2005)

Chemist and champion of nanotechnology.

Towards the end of his life, Richard Smalley had begun to say, "If it ain't tubes, we don't do it". He had become fascinated by the prospect that fullerenes — the massive carbon molecules with distinctive geometrical shapes that he had co-discovered in 1985 — might be re-formed into single-walled nanotubes with exciting properties. In particular, he dreamt of making a metallically conducting cable of billions of these carbon nanotubes, which, for the same weight, would be many times stronger than steel. Smalley's time ran out before he achieved that goal; nevertheless, the legacy of this research already extends far beyond the confines of materials science, to such diverse fields as energy technology and medicine.

The single-minded obsession that Smalley, who died on 28 October, brought to nanotube research was in fact rather out of character. In his early career as an independent researcher, he had tended to create a new research field about every two years, often abandoning them with equal frequency. The tenor of this period was set in postdoctoral work at the University of Chicago, where he pioneered a technique that combined laser excitation of molecules with their cooling by supersonic jets of gas. He demonstrated that the method greatly simplified the complex spectra of the molecules' energy levels, and allowed complexes bound together by very weak van der Waals interactions to be created and observed. On arriving, in 1976, at Rice University in Houston, Texas — where he was to stay for the rest of his career — he rapidly created a series of spectroscopic tools based on this technique that are used to this day. Smalley's approach was to conceive a way to investigate a chemical system or phenomenon, construct the necessary sophisticated apparatus, do enough work to show the true potential of the method, and move on. Each new project was better than the last, offering further valuable scientific information.

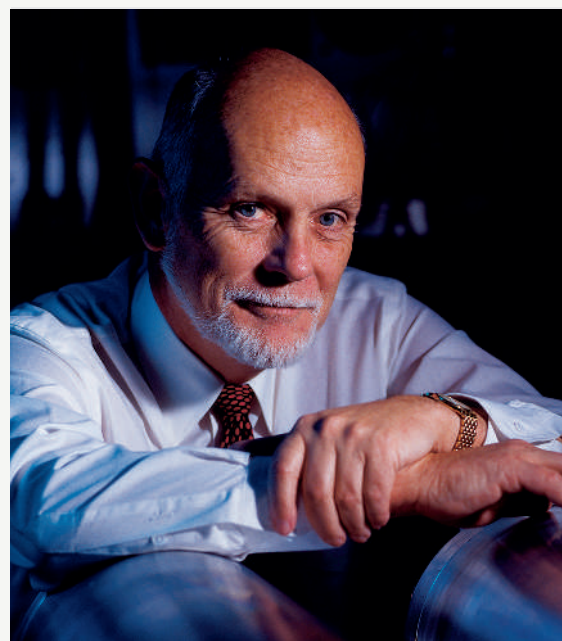
The discovery of the fullerenes, which led to his Nobel Prize in Chemistry in 1996, grew out of one such project. In this, Smalley was studying jet-cooled molecular clusters formed by the condensation of laser-vaporized metals or semiconductors. In March 1984, the British chemist Harry Kroto, whose radio astronomy observations had detected long carbon-chain compounds in interstellar clouds, visited Rice. Kroto saw the vaporizing graphite in Smalley's apparatus as a way of testing his idea that these chains

were being formed by the condensation of species ejected from carbon-rich stars. When the experiments were finally performed at the facility in September 1985, proof for the formation of carbon chains between 7 and 12 atoms long, the size range of the astronomical observations, was indeed found.

The experiments also showed striking evidence that more interesting, much larger carbon clusters of between 40 and 80 atoms were being formed simultaneously. The particularly high abundance of the C_{60} cluster could only be explained if it were a stable, closed cage with 20 hexagonal and 12 pentagonal interlocking faces, rather like a football, or soccer ball. Because of the transatlantic conflict between the experimentalists over the exact name of the ball and the sport it belonged to — and because the structure was reminiscent of the geodesic domes of the architect Buckminster Fuller — the C_{60} structure was named buckminsterfullerene. A more general examination of the number of carbon atoms in the other, differently sized clusters that were found led to the gradual realization that they must all be carbon cages consisting of exactly 12 pentagons and a number of hexagons that grew with increasing cluster size.

Efforts in Smalley's laboratory to make a macroscopic sample of buckminsterfullerene were abandoned fairly quickly after experiments to vaporize a graphite rod using a laser left no trace of C_{60} . The isolation in 1990 of a mixture of C_{60} and C_{70} , using an apparatus consisting of a carbon arc inside a bell jar, seemed ridiculously simple compared with Smalley's high-tech approach. Smalley reinvestigated the laser vaporization technique, and found that the amount of C_{60} produced depended strongly on the temperature of the wall of the quartz tube that surrounded the graphite rod: no C_{60} was obtained when it was at room temperature, but there was a 20% yield at 1,100 °C.

The isolation of single-walled carbon nanotubes (SWNTs), announced in June 1993, soon drove Smalley's attention and considerable powers to another domain. The development of synthesis techniques for SWNTs was a challenge unlike anything Smalley had encountered before 1990, and virtually all that was known was the need for a metal catalyst — iron, cobalt or nickel. Soon Smalley found that, by impregnating these metals into the graphite rods used to make C_{60} in the laser vaporization experiments, he



could create SWNTs in the form of ropes containing more than a hundred individual tubes. Between 1993 and 2005, Smalley found a generally better way of making the tubes, as well as ways of cutting them up, performing chemical reactions on them and producing them in solution. In the last week of his life, desperately ill with leukaemia, he was enthusiastically receiving progress reports in his hospital bed and suggesting new ideas and experiments.

Rick Smalley was a remarkable person, both professionally and personally. He had two sons almost thirty years different in age, and four wives — the first two of whom were his guests at the Nobel ceremony in 1996 (when he himself was single). Everyone got along amiably, both former wives seeming to have a wonderful time. As his end neared, Smalley's fourth wife Deborah and older son Chad cared for him constantly.

Smalley's ability to vacuum up information, organize it and use it for creative scientific endeavour was prodigious. He always tackled the most challenging problems, was indefatigable in the pursuit of answers, and in all arguments met logic with logic. Smalley had a whimsical sense of humour and tremendous personal charisma. Others usually found it to their advantage to follow his lead, as collaboration with Smalley generally resulted in excellent scientific results. Smalley's persuasiveness came most effectively to bear in the campaign to convince the US government to create its National Nanotechnology Initiative, a great achievement in public policy. ■

Robert F. Curl

Robert F. Curl is in the Department of Chemistry, Rice University, Houston, Texas 77251-1892, USA. He shared the 1996 Nobel Prize in Chemistry with Richard Smalley and Harry Kroto.

T. LAVERGNE/RICE UNIV.

BRIEF COMMUNICATIONS

Circadian organization in reindeer

These Arctic animals abandon their daily rhythms when it is dark all day or light all night.

The light/dark cycle of day and night synchronizes an internal 'biological clock' that governs daily rhythms in behaviour, but this form of regulation is denied to polar animals for most of the year. Here we demonstrate that the continuous lighting conditions of summer and of winter at high latitudes cause a loss in daily rhythmic activity in reindeer living far above the Arctic Circle. This seasonal absence of circadian rhythmicity may be a ubiquitous trait among resident polar vertebrates.

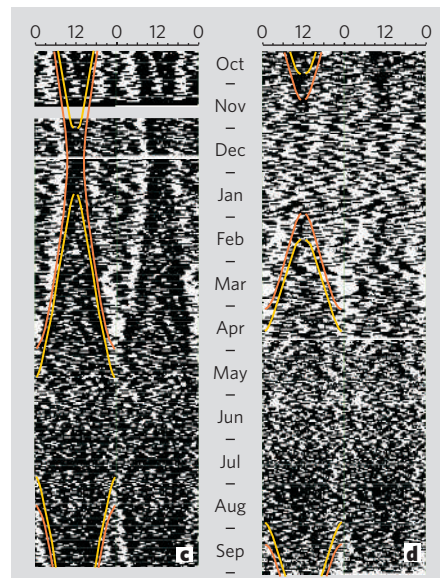
Circadian oscillators are present in all organisms on our rotating planet (see ref. 1, for example). These 'biological clocks' govern the temporal organization of physiological functions and behaviour, and enable plants and animals to anticipate and prepare for daily events such as sunrise and sunset². A convincing argument for the internal nature of circadian regulation is the persistence of temporal organization under constant environmental conditions. This was first described in the plant *Mimosa pudica* by the French astronomer Jean de Mairan in 1729 (ref. 3), and persistence, under constant conditions, of rhythms with periods deviating slightly from 24 h remains a prerequisite for the identification of circadian control.

Few plants and animals experience constant conditions in the wild. But at high latitudes, where the Sun neither sets in summer nor rises in winter, resident polar organisms experience the distinct changes in light intensity that result from a day/night cycle for just a few weeks each year — in spring and in autumn. We recorded daily patterns of activity continuously for one year in two subspecies of reindeer: *Rangifer tarandus platyrhynchus* (5 or 6 animals; Fig. 1a, b) living at 78° N in the high Arctic archipelago of Svalbard, and *R. t. tarandus* (6 animals) living in northern Norway at 70° N. The animals ranged freely in their natural habitat. (For methods, see supplementary information.)

All animals showed alternating bouts of activity and inactivity typical of ruminants⁴; these cycles were substantially shorter than 24 hours (ultradian) and persisted throughout the year. Plots of activity over time (actograms) reveal a complete absence of circadian organization of this behaviour in both subspecies in summer and in Svalbard reindeer in winter (Fig. 1c, d). Evidently, the changes in light intensity that occur across the day at these times are not sufficient to synchronize



Figure 1 | Activity patterns in reindeer under polar light conditions. a, b, Polar light conditions: mixed groups of Svalbard reindeer *Rangifer tarandus platyrhynchus* at 78° N, at a, midnight in late June, and b, midday in mid-February. c, d, Sample actograms showing patterns of activity over one year in sub-adult reindeer in c, northern Norway (*R. t. tarandus*, 70° N; $n = 1$), and d, Svalbard (*R. t. platyrhynchus*, 78° N; $n = 1$). Data, recorded continuously using small activity-loggers, are presented as double-plot actograms in which each row represents two consecutive days; time of day is indicated. Bouts of activity (black bars) are interspersed with bouts of inactivity (white spaces). Grey region, data missing. Lines indicating the beginning and end of civil twilight (when light intensity is 10 lux, orange) and sunrise and sunset (yellow) are superimposed on each actogram. Rhythmicity in the actograms was determined by F-periodogram analysis (see supplementary information).



the pattern of activity in reindeer. However, the daily pattern was modified when there was a distinct light/dark cycle, and reindeer in northern Norway, in particular, displayed a significant rhythm of exactly 24 hours throughout autumn, winter and spring (see supplementary information).

Free-living reindeer do not therefore show evidence of the classical prerequisite for circadian organization—persistence under constant conditions. Seasonal absence of rhythmicity in the circadian range has been recorded in the daily activity of the Svalbard ptarmigan (*Lagopus mutus hyperboreus*)⁵, and in circulating levels of the hormone melatonin in ptarmigan⁶ and in reindeer⁷, indicating that the expression of an internal clock is reduced in both Arctic species under constant light conditions. We therefore suggest that seasonal

absence of circadian rhythmicity is a ubiquitous trait among resident polar vertebrates.

Reduced circadian organization may enhance animals' responsiveness and speed of phase adaptation to the light/dark cycle, as proposed for migrating birds⁸ and mammals emerging from hibernation⁹. And for herbivores in polar regions, there can be little selective advantage in maintaining strong internal clocks in an effectively non-rhythmic environment.

Bob E. H. van Oort*, **Nicholas J. C. Tyler†**,
Menno P. Gerkema‡, **Lars Folkow***,
Arnoldus Schytte Blix*, **Karl-Arne Stokkan***

*Department of Arctic Biology and Institute of Medical Biology, †Centre for Sámi Studies, University of Tromsø, 9037 Tromsø, Norway
e-mail: karlarne@fagmed.uit.no

‡Department of Chronobiology, University of

Groningen, 9750 AA Haren, The Netherlands

1. Lowrey, P. L. & Takahashi, J. S. *Annu. Rev. Genom. Hum. Genet.* **4**, 407–441 (2004).
2. Daan, S. & Aschoff, J. in *Circadian Clocks* (eds Takahashi, J. S. et al.) 7–43 (Plenum, New York, 2001).
3. de Mairan, J. J. in *Histoire de l'Académie Royale des Sciences* 35–36 (Paris, 1729).
4. Gerkema, M. P. in *Biological Rhythms* (ed. Kumar, V.) 207–215 (Narosa, New Delhi, 2002).
5. Stokkan, K. A., Mortensen, A. & Blix, A. S. *Am. J. Physiol.* **251**, 264–267 (1986).
6. Reiherth, E., van't Hof, T. & Stokkan, K. A. *J. Biol. Rhyth.* **14**, 314–319 (1999).
7. Stokkan, K. A., Tyler, N. J. C. & Reiter, R. J. *Can. J. Zool.* **72**, 904–909 (1994).
8. Hau, M. & Gwinner, E. *Physiol. Behav.* **58**, 89–95 (1995).
9. Hut, R. A., Van der Zee, E. A., Jansen, K., Gerkema, M. P. & Daan, S. *J. Comp. Physiol. B* **172**, 59–70 (2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/4381095a

WORLD YEAR OF PHYSICS

A direct test of $E = mc^2$

One of the most striking predictions of Einstein's special theory of relativity is also perhaps the best known formula in all of science: $E = mc^2$. If this equation were found to be even slightly incorrect, the impact would be enormous — given the degree to which special relativity is woven into the theoretical fabric of modern physics and into everyday applications such as global positioning systems. Here we test this mass–energy relationship directly by combining very accurate measurements of atomic-mass difference, Δm , and of γ -ray wavelengths to determine E , the nuclear binding energy, for isotopes of silicon and sulphur. Einstein's relationship is separately confirmed

in two tests, which yield a combined result of $1 - \Delta mc^2/E = (-1.4 \pm 4.4) \times 10^{-7}$, indicating that it holds to a level of at least 0.00004%. To our knowledge, this is the most precise direct test of the famous equation yet described.

Our direct test is based on the prediction that when a nucleus captures a neutron and emits a γ -ray, the mass difference Δm between the initial (including unbound neutron) and final nuclear states, multiplied by c^2 (where c is the speed of light), should equal the energy of the emitted γ -ray(s), as determined from Planck's relation $E = hf$ (where h is Planck's constant and f is frequency).

The total energy of the γ -rays emitted as the daughter nucleus decays to the ground-state was determined by summing the individual γ -ray energies. These were obtained by wavelength measurement using crystal Bragg spectroscopy. The mass difference Δm is measured by simultaneous comparisons of the cyclotron frequencies (inversely proportional to the mass) of ions of the initial and final isotopes confined over a period of weeks in a Penning trap.

For an atom X with a mass number of A

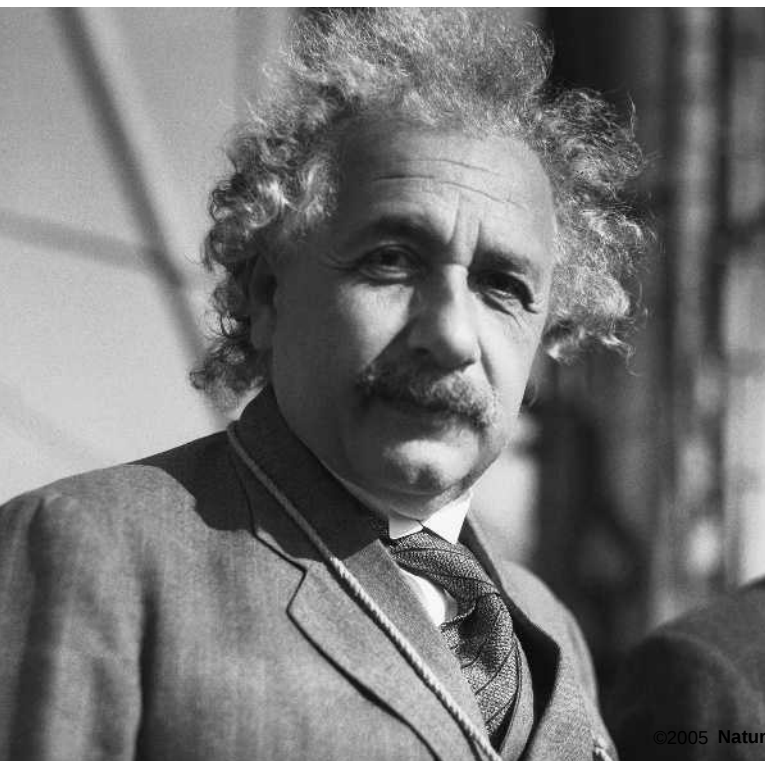
undergoing this nuclear reaction, the comparison is expressed in terms of measured quantities as

$$\Delta Mc^2 = (M[^A\text{X}] - M[^{A+1}\text{X}] + M[\text{D}] - M[\text{H}])c^2 = 10^3 N_A h (f_{A+1} - f_D) \text{ mol AMU kg}^{-1} \quad (1)$$

where the Avogadro constant N_A relates the measured mass $M[\text{X}]$ in unified atomic mass units (AMU) to its mass in kilograms $m[\text{X}]$. We made comparisons for $^{A+1}\text{X} = ^{29}\text{Si}$ and $^{A+1}\text{X} = ^{33}\text{S}$. The mass of the neutron $M[n]$ is determined from the masses¹ of hydrogen $M[\text{H}]$ and deuterium $M[\text{D}]$ combined with f_D , the frequency of the γ -ray corresponding to the deuterium binding energy². The molar Planck constant is $N_A h = 3.990312716(27) \times 10^{-10} \text{ J s mol}^{-1}$; numbers in parentheses indicate uncertainty on the last digits. This figure has been independently confirmed at about the 5×10^{-8} level by a range of experiments through its relationship with the fine-structure constant¹.

The γ -ray frequencies on the righthand side of equation (1) have been measured using the GAMS4 crystal-diffraction facility at the Laue–Langevin Institute in Grenoble³. The γ -rays emitted from sources located near the high-flux reactor core are diffracted by two nearly perfect, flat crystals whose lattice spacing, d , has been determined in metres⁴. The diffraction angles, θ , are measured with angle interferometers calibrated using a precision optical polygon (Fig. 1a). Wavelengths are determined from the Bragg equation $\lambda = 2d \sin \theta$ and were measured for the 3.5-MeV and 4.9-MeV transitions in ^{29}Si , for the 0.8-MeV, 2.4-MeV and 5.4-MeV transitions in ^{33}S , and for the 2.2-MeV transition in deuterium ^2H (see supplementary information).

Because the diffraction angle for a 5-MeV γ -ray by a low-order silicon reflection is less than 0.1° , our binding-energy determinations were limited by our ability to measure the diffraction angles of the highest-energy γ -rays with fractional accuracy better than about



BETTAMIN/CORBIS

Albert Einstein: father of the famous formula.

10^{-7} . A slightly revised value of f_D from that given in ref. 2 is also reported here (c is exact):
 $f_{Si} = c/(0.146318275(86) \times 10^{-12} \text{ m})$
 $f_S = c/(0.143472991(54) \times 10^{-12} \text{ m})$
 $f_D = c/(0.557341007(98) \times 10^{-12} \text{ m})$

These values of f_{Si} and f_S are, respectively, 25 and 50 times more precise than earlier values⁴.

The mass difference was determined by direct comparison of the cyclotron frequencies of two different ions simultaneously confined in a Penning trap⁵. This two-ion technique achieves mass comparisons with fractional accuracies below 10^{-11} by virtually eliminating the effect of many sources of noise such as magnetic field fluctuations. During measurements, the two ions are placed on a common circular orbit (magnetron mode), on opposite sides of the centre of the trap and separated by a distance of about 1 mm. Figure 1b shows that the measured cyclotron frequency ratio is almost completely independent of ion-ion separation, tightly constraining the size of the largest possible systematic errors⁵. The ion mass ratios reported here, namely $M[^{33}\text{S}^+]/M[^{32}\text{SH}^+] = 0.9997441643450(89)$ and $M[^{29}\text{Si}^+]/M[^{28}\text{SiH}^+] = 0.9997151241812(65)$, are respectively more than 700 and 100 times more precise than previously known⁶.

The total uncertainties include the uncorrelated uncertainties of 4×10^{-12} each from ion-ion interactions, trap field imperfections, and statistics. The ratios also include corrections of $45(5) \times 10^{-12}$ and $7.3(2) \times 10^{-12}$ to account for polarization-induced shifts of the cyclotron frequencies⁷ (see methods in supplementary information). After accounting for the mass of the missing electron and the chemical binding energies^{8,9}, the ion mass ratios give the neutral mass differences: $M[^{32}\text{S}] + M[\text{H}] - M[^{33}\text{S}] = 0.00843729682(30) \text{ AMU}$, and $M[^{28}\text{Si}] + M[\text{H}] - M[^{29}\text{Si}] = 0.00825690198(24) \text{ AMU}$. And by the addition of $M[\text{D}] - 2M[\text{H}] = -0.00154828629(40) \text{ AMU}$ (refs 1,6) to each one, we obtain the desired mass differences of equation (1), with a fractional accuracy of about 7×10^{-8} for both.

Comparing the measured energies and

Figure 1 | Typical data for testing $E = \Delta mc^2$. **a**, ^{33}S wavelength data and fitted curves. The energy of the emitted 5,421-keV γ -rays is measured from the angular separation of the second-order Bragg peaks resulting from diffraction from a silicon crystal. **b**, Measured mass ratio $M[^{33}\text{S}^+]/M[^{32}\text{SH}^+]$ (which largely determines the mass change, Δm) as a function of the distance between the ions, ρ , in the Penning trap. The lack of variation strongly suggests that the predicted uncertainties from ion-ion interactions and field inhomogeneities (turquoise and blue-hatched bands, respectively) are too conservative. The final mass ratio (solid line) is determined with fractional accuracy of 7 parts in 10^{12} (dashed lines). All error bars are one standard deviation.

masses, we report two independent tests of $E = mc^2$. The two comparisons find a measured fractional difference between E and Δmc^2 of $2.1(5.2) \times 10^{-7}$ and $-9.7(8.0) \times 10^{-7}$ with sulphur and silicon isotopes, respectively, and a combined value of $-1.4(4.4) \times 10^{-7}$. The error on this comparison is currently dominated by the uncertainty on the γ -ray measurements. This result is 55 times more accurate than the previous best direct test of $E = mc^2$, which was performed by comparing the electron and positron masses to the energy released in their annihilation¹⁰.

Simon Rainville*†, **James K. Thompson***,
Edmund G. Myers‡, **John M. Brown§**,
Maynard S. Dewey||, **Ernest G. Kessler Jr||**,
Richard D. Deslattes||, **Hans G. Börner¶**,
Michael Jentschel¶, **Paolo Mutti¶**,
David E. Pritchard*

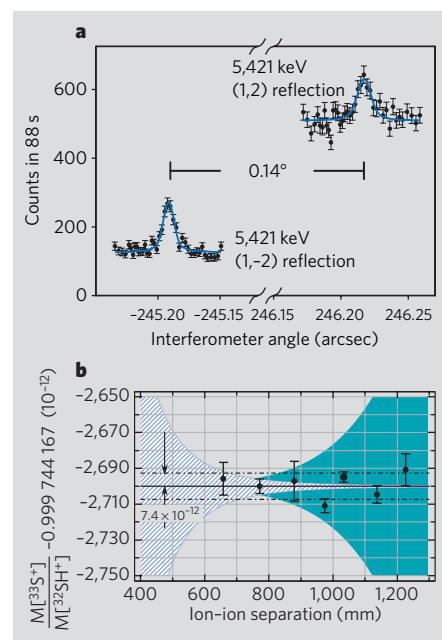
*Research Laboratory of Electronics, MIT-Harvard Center for Ultracold Atoms, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

e-mail: rainville@alum.mit.edu

†Present address: Département de Physique, Université Laval, Québec G1K 7P4, Canada

‡Department of Physics, Florida State University, Tallahassee, Florida 32306-4350, USA

§The Physical and Theoretical Chemistry Laboratory, Department of Chemistry, South



Parks Road, Oxford OX1 3QZ, UK

||National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

¶Institut Laue-Langevin, 38042 Grenoble Cedex, France

- Mohr, P. J. & Taylor, B. N. *Rev. Mod. Phys.* **77**, 1–107 (2005).
- Kessler, E. G. *et al. Phys. Lett. A* **255**, 221–229 (1999).
- Kessler, E. G. *et al. Nucl. Instrum. Meth. Phys. Res. A* **457**, 187–202 (2001).
- Dewey, M. S. *et al.* Preprint nucl-ex/0507011 at <http://arxiv.org> (2005).
- Rainville, S., Thompson, J. K. & Pritchard, D. E. *Science* **303**, 334–338 (2004).
- Audi, G., Wapstra, A. H. & Thibault, C. *Nucl. Phys. A* **729**, 337–676 (2003).
- Thompson, J. K., Rainville, S. & Pritchard, D. E. *Nature* **430**, 58–61 (2004).
- NIST Chemistry WebBook, NIST Standard Reference Database Number 69 (eds Linstrom, P. J. & Mallard, W. G.) (National Institute of Standards and Technology, Gaithersburg, Maryland, 2003).
- Chase, M. W. *J. Phys. Chem. Ref. Data Monogr.* **9**, 1–1951 (1998).
- Greene, G. L., Dewey, M. S., Kessler, E. G. & Fischbach, E. *Phys. Rev. D* **44**, R2216–R2219 (1991).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/4381096a

CHEMICAL COMMUNICATION

Chirality in elephant pheromones

Musth in male elephants is an annual period of heightened sexual activity and aggression^{1–3} that is linked to physical, sexual and social maturation. It is mediated by the release of chemical signals such as the pheromone frontalin, which exists in two chiral forms (molecular mirror images, or enantiomers). Here we show that enantiomers of frontalin are released by Asian elephants in a specific ratio that depends on the animal's age and stage of musth, and that different responses are elicited in male and female conspecifics

when the ratio alters. This precise control of communication by molecular chirality offers insight into societal interactions in elephants, and may be useful in implementing new conservation protocols⁴.

Frontalin (1,5-dimethyl-6,8-dioxabicyclo [3.2.1]octane)^{1,2,5,6} is discharged during musth in male Asian elephants (*Elephas maximus*) from the temporal gland on the face. We analysed more than 100 secretion samples from six males and found that this pheromone was first detectable in the late teens, the quan-

tity secreted rising about 15-fold over a 25-year age span (Fig. 1a). Both enantiomeric forms (designated plus and minus; Fig. 1b, inset) were present and each was quantified (for methods, see supplementary information).

In young males, significantly more (+) than (−) frontalin was secreted, but concentrations of these became almost equal (1:1, a racemic mixture) as the elephant matured (Fig. 1b). Serial samples collected throughout the musth episodes of two young and two old males confirmed that this change in ratio was significantly associated with the elephants' age groups (results not shown).

Musth periods get longer as males age. The enantiomeric composition of the frontalin secreted becomes significantly more varied in

the longer musths of older males but, critically, the ratios are nearly racemic during mid-musth — the period of prime signalling (Fig. 1c, right). Wide fluctuations in the ratio of frontalinal enantiomers from young males (Fig. 1c, left) are akin to the variations in serum androgens that occur during the progression of their shorter 'moda', or honey musth¹.

To test whether the enantiomeric composition of frontalinal affects its activity as a multi-purpose pheromone (attractant, repellent, deterrent and dominance enhancer), we investigated the effects on five categories of conspecific (females: follicular, luteal-phase or pregnant; males: young or old). Total frontalinal concentrations and frontalinal enantiomer ratios were determined for samples of mid-musth, temporal-gland secretions from males with a range of ages; the responses of other conspecifics to these samples were then recorded — behaviour such as sniffing and checking indicated attraction, whereas responses such as running away and circling the sample indicated repulsion.

Other elephants were generally indifferent to secretions containing no frontalinal, apart from young males (Fig. 1d, left). Secretions with low concentrations of frontalinal, predominantly as the (+) enantiomer, aroused the interest of young males and were of mild interest to old males (Fig. 1d, middle). Secretions containing high concentrations of frontalinal at racemic ratios actively repulsed males of all ages, as well as luteal-phase and pregnant females, whereas they attracted follicular-phase females (Fig. 1d, right).

Our results indicate that the ratio of frontalinal enantiomers enables other elephants to distinguish both the maturity of male elephants in musth and the phase of musth. They corroborate earlier findings that synthetic racemic frontalinal can elicit different behavioural responses from conspecifics, depending on their sex, age and reproductive status².

The length of musth increases as males mature, with the fittest sustaining a long mid-musth phase and releasing an optimal ratio of frontalinal enantiomers, thereby improving their status and enjoying easier access to females³; these mid-musth emissions are picked up almost exclusively by ovulating females. A similar quantitative blending of frontalinal enantiomers is also necessary to elicit

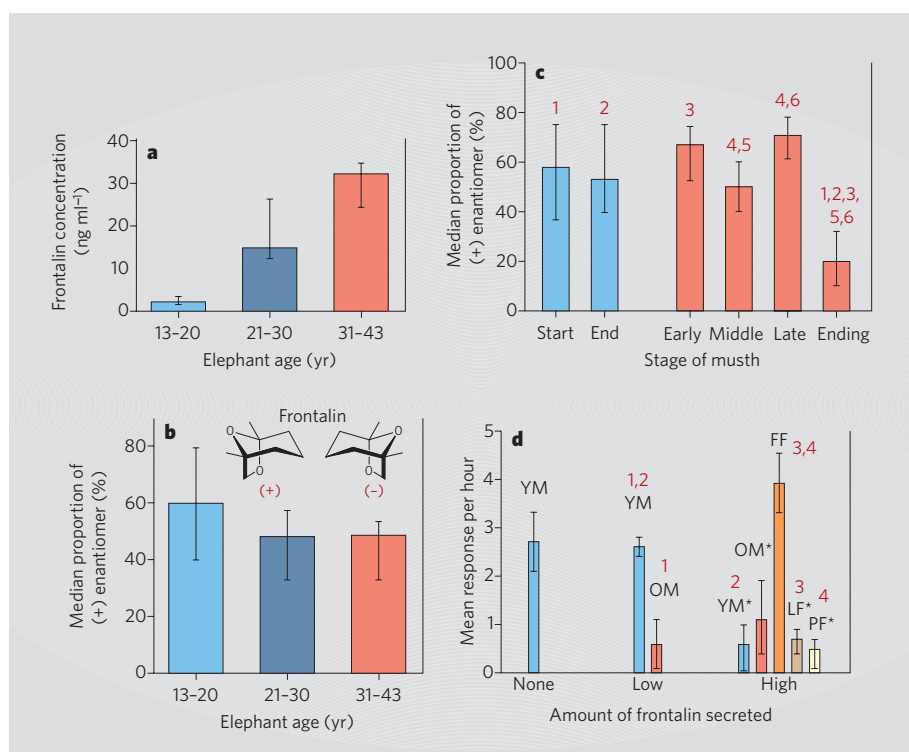


Figure 1 | Secretion of the pheromone frontalinal by male Asian elephants in musth. **a**, Median concentrations of frontalinal from elephants in three age groups (number of animals in each was 37, 41 and 14, respectively; statistically significant differences (s.s.d.) between groups, $P < 0.05$). **b**, Changing median proportions of (+) and (−) enantiomers of frontalinal (inset) with increasing male age (numbers in each age group were 31, 48 and 16, respectively); s.s.d., $P < 0.05$, as indicated by paired numbers. **c**, Changes in median proportion of (+) enantiomer during musth progression for young (blue) and old males (red); s.s.d., $P < 0.05$, between paired numbers. **d**, Mean responses by conspecifics to different frontalinal concentrations; s.s.d. (t -test), $P = 0.001$, between paired numbers. YM, young males; OM, older males; FF, follicular females; LF, luteal females; PF, pregnant females; asterisks, avoidance of musth male; no asterisk, attraction to musth male. Secretions with no frontalinal induced no response in older males and a very low response in females; females were not tested against low frontalinal concentrations. For details of methods, see supplementary information.

a full biological response in *Dendroctonus* beetles^{7–12}. Our discovery of a stereospecifically tailored message in mammals should help in resolving the interactions of pheromones with their respective receptor proteins, which are also chiral.

David R. Greenwood*, **Dan Comeskey†**, **Martin B. Hunt†**, **L. Elizabeth L. Rasmussen‡**

*HortResearch, Mount Albert Research Centre, Private Bag 92-169, Auckland, New Zealand

†Palmerston North Research Centre, Hort Research, Private Bag 11-030, Palmerston North, New Zealand

‡Department of Environmental and Biomolecular Systems, OGI School of Science and Engineering, Oregon Health and Sciences University, Beaverton, Oregon 97006, USA
e-mail: betsr@ebs.ogi.edu.

- Rasmussen, L. E. L., Riddle, H. S. & Krishnamurthy, V. *Nature* **415**, 975–976 (2002).
- Rasmussen, L. E. L. & Greenwood, D. R. *Chem. Senses* **28**, 433–446 (2003).
- Rasmussen, L. E. L., Krishnamurthy, V. & Sukumar, R. *Behaviour* **142**, 351–396 (2005).
- Rasmussen, L. E. L. & Riddle, S. W. *J. Elephant Man. Assoc.* **15**, 30–38 (2004).
- Rasmussen, L. E. L. & Perrin, T. E. *Physiol. Behav.* **67**, 539–549 (1999).
- Perrin, T. E., Rasmussen, L. E. L., Gunawardena, R. & Rasmussen, R. A. *J. Chem. Ecol.* **21**, 207–221 (1996).
- Silverstein, R. M. *J. Chem. Ecol.* **14**, 1981–2003 (1988).
- Mori, K. *Chirality* **10**, 578–585 (1998).
- Paine, T. D., Millar, J. G., Hanlon, C. C. & Hwang, J. S. *J. Chem. Ecol.* **25**, 433–453 (1999).
- Kinzer, G. W. *et al. Nature* **221**, 477–478 (1969).
- Francke, W. *et al. J. Chem. Ecol.* **21**, 1043–1063 (1995).
- Stewart, T. E., Plummer, E. L., McCandless, L. L., West, J. R. & Silverstein, R. M. *J. Chem. Ecol.* **3**, 27–43 (1977).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/4381097a



A musth have: mature Asian elephants (*Elephas maximus*) secrete a chiral chemical in a ratio that is especially attractive to females.

METEOROLOGY

Are there trends in hurricane destruction?

Arising from: K. Emanuel *Nature* **436**, 686–688 (2005)

Since the record impact of Hurricane Katrina, attention has focused on understanding trends in hurricanes and their destructive potential. Emanuel¹ reports a marked increase in the potential destructiveness of hurricanes based on identification of a trend in an accumulated annual index of power dissipation in the North Atlantic and western North Pacific since the 1970s. If hurricanes are indeed becoming more destructive over time, then this trend should manifest itself in more destruction. However, my analysis of a long-term data set of hurricane losses in the United States shows no upward trend once the data are normalized to remove the effects of societal changes.

Historical hurricane losses can be adjusted to a base year's values through adjustments related to inflation, population and wealth². For at least three reasons, this data set is appropriate for identifying long-term climate signals. First, a long-term record of flood damage (collected in a similar way to and by the same agency as the hurricane data) is of sufficient quality to identify long-term trends³. Second, a methodology² developed in 1998 produces results that are consistent with the results of catastrophe models used by the insurance industry to assess hurricane losses⁴. Third, and most crucially, the data set contains climate signals, such as that of the El Niño–Southern Oscillation, which has a well established climatological relationship with interannual hurricane behaviour (see refs 5, 6, for example).

Specifically, an index of sea-surface-temperature anomalies of the Niño 3.4 region of the central Pacific in August, September and October is highly correlated with observed normalized damages in the same year⁵. The observed intensity change⁷ in Atlantic basin hurricanes between El Niño and La Niña events is of similar magnitude to the changes in annual accumulated power-dissipation index identified by Emanuel¹; the ability to identify the signal of the former suggests therefore that the normalized damage database is of sufficient size and quality to identify climate signals of the magnitude discussed by Emanuel.

A data set of hurricane losses (focusing on direct damages related to wind, and generally excluding rain-caused flood damage) for individual storms⁶ extended to 2004, which includes only those storms causing damage, shows no upward trend. For example, take the 86 storms causing at least US\$1 billion in normalized damages, which removes a bias caused by small storms resulting in no damage in the early twentieth century (that is, not subjected to normalization). There is an average per-storm loss in 1900–50 for 40 storms (0.78

events per year) of \$9.3 billion, and an average per-storm loss in 1951–2004 for 46 storms (0.85 events per year) of \$7.0 billion; this difference is not statistically significant. Adding Hurricane Katrina to this data set, even at the largest loss figures currently suggested, would not change the interpretation of these results.

These loss data indicate two possibilities with respect to Emanuel's analysis¹: if the power-dissipation index metric is an accurate indicator of hurricane destructiveness, then the trend identified by Emanuel could be an artefact of the data and/or methods; alternatively, the trend he identifies is an accurate reflection of trends in the real-world characteristics of storms, but the power-dissipation index is a weak indicator of hurricane destructiveness — which would call for the identification of climate metrics more directly associated with societal outcomes. In any case, it is misleading to characterize Emanuel's results as indicating an increase in “destructiveness” or as an indication of future increases in destruction resulting from changes in the power-dissipation index.

The bottom line is that, with no long-term trend identified in normalized hurricane damage over the twentieth century (in the United

States or elsewhere; see ref. 8, for example), it is exceedingly unlikely that scientists will identify large changes in historical storm behaviour that have significant societal implications. Looking to the future, Emanuel¹ provides no evidence to alter the conclusion that changes in society will continue to have a much larger effect than changes in climate on the escalating damage resulting from tropical cyclones⁹.

Roger A. Pielke, Jr

Center for Science and Technology Policy Research, University of Colorado, Campus Box 488, Boulder, Colorado 80309-0488, USA
e-mail: pielke@cires.colorado.edu

1. Emanuel, K. *Nature* **436**, 686–688 (2005).
2. Pielke, R. A. Jr & Landsea, C. W. *Bull. Am. Meteorol. Soc.* **13**, 621–631 (1998).
3. Downton, M. & Pielke, R. A. Jr *Natural Hazards* **35**, 211–228 (2005).
4. Pielke, R. A. Jr, Landsea, C. W., Downton, M. & Muslin, R. *J. Insur. Reg.* **18**, 177–194 (1999).
5. Katz, R. W. J. *J. Appl. Meteorol.* **41**, 754–762 (2002).
6. Pielke, R. A. Jr & Landsea, C. W. *Bull. Am. Meteorol. Soc.* **80**, 2027–2033 (1999).
7. Landsea, C. L., Pielke, R. A. Jr, Mestas-Núñez, A. & Knaff, J. *Clim. Change* **42**, 89–129 (1999).
8. Raghavan, S. & Rajesh, S. *Bull. Am. Meteorol. Soc.* **84**, 635–644 (2003).
9. Pielke, R. A. Jr et al. *Bull. Am. Meteorol. Soc.* **86**, 1571–1575 (2005).

doi:10.1038/nature04426

METEOROLOGY

Hurricanes and global warming

Arising from: K. Emanuel *Nature* **436**, 686–688 (2005)

Anthropogenic climate change has the potential for slightly increasing the intensity of tropical cyclones through warming of sea surface temperatures¹. Emanuel² has shown a striking and surprising association between sea surface temperatures and destructiveness by tropical cyclones in the Atlantic and western North Pacific basins. However, I question his analysis on the following grounds: it does not properly represent the observations described; the use of his Atlantic bias-removal scheme may not be warranted; and further investigation of a substantially longer time series for tropical cyclones affecting the continental United States does not show a tendency for increasing destructiveness. These factors indicate that instead of “unprecedented” tropical cyclone activity having occurred in recent years, hurricane intensity was equal or even greater during the last active period in the mid-twentieth century.

My first concern is that Emanuel's figures²

do not match their description: his Figs 1–3 aim to present smoothed power-dissipation index (PDI) time series with two passes of a 1–2–1 filter, but the end-points — which are crucial to his conclusions — instead retain data unaltered by the smoothing; this is important because the last data point plotted in Emanuel's Fig. 1 is far larger than any other portion of the time series. Even after adding last year's busy hurricane season into the analysis and then properly using the filter, as described, the crucial end-point of the smoothed time series no longer jumps up dramatically in the last couple of years (Fig. 1a). About one-third of the increase in Atlantic PDI in Emanuel's graph for the past ten years is incorrect owing to inappropriate plotting of the data, even if the active 2004 season is incorporated.

A second concern is the bias-removal scheme used to alter the data for the Atlantic for 1949–69. Emanuel can demonstrate

“unprecedented” activity in the past ten years only by markedly reducing the tropical-cyclone winds for the first two decades of the time series. He attempts to use a bias-removal scheme³ that recommends reduction of the tropical-cyclone winds by $2.5\text{--}5.0\text{ m s}^{-1}$ for the 1940s–60s because of an inconsistency in the pressure–wind relationship during those years compared with subsequent (and presumably more accurate) data. However, the

function used by Emanuel to reduce the winds in the earlier period goes well beyond this recommendation, as the bias removal used continued to increase with increasing wind intensity and reached a reduction of as much as 12.2 m s^{-1} for the strongest hurricane in the 1949–69 original data set.

In major hurricanes, winds are substantially stronger at the ocean’s surface^{4–7} than was previously realized, so it is no longer clear that

Atlantic tropical cyclones of the 1940s–60s call for a sizeable systematic reduction in their wind speeds. It is now understood to be physically reasonable that the intensity of hurricanes in the 1970s through to the early 1990s was underestimated, rather than the 1940s and 1960s being overestimated⁸. To examine changes in intensity over time, it is therefore better to use the original hurricane database than to apply a general adjustment to the data in an attempt to make it homogenous.

Figure 1b shows Emanuel’s bias-removed smoothed curve and the substantially larger PDI values in the original hurricane data set; the latter indicates that amplitudes for 1949–69 are comparable to those for the most recent decade. This is consistent with earlier work^{9,10}, emphasizing the large multidecadal oscillations in activity. It is also likely that values of PDI from the 1940s to the mid-1960s are substantially undercounted owing to the lack of routine aircraft reconnaissance and geostationary satellite monitoring of tropical cyclones far from land.

A third concern is that it is difficult to separate out any anthropogenic signal from the substantial natural multidecadal oscillations with a relatively short record of tropical-cyclone activity. One way to extend the PDI analysis back to include several additional decades of reliable records is to examine only those tropical cyclones that made landfall along populated coastlines^{11,12}. Figure 2 shows that tropical-cyclone activity in the United States was generally extremely busy between the 1930s and 1960s, but fell below average between the 1970s and early 1990s. Despite the extreme value for 2004, the most recent decade has a PDI that is near-average for the United States, rather than showing an increase in the overall number and intensity of hurricane strikes.

Despite these problems, Emanuel’s study illustrates the pressing need for a completion of the storm-by-storm reanalysis of the Atlantic hurricane database^{8,11}, which will provide a more homogeneous time series of tropical-cyclone intensities and so avoid the application of arbitrary bias-removal schemes. But, on the basis of the evidence I present here, claims to connect Atlantic hurricanes with global warming are premature. The Atlantic hurricane basin is currently seeing enhanced, rather than “unprecedented”, storminess that is comparable to, or even less active than, that seen in earlier busy cycles of activity.

Christopher W. Landsea

NOAA/AOML/Hurricane Research Division,
Miami, Florida 33149, USA
e-mail: chris.landsea@noaa.gov

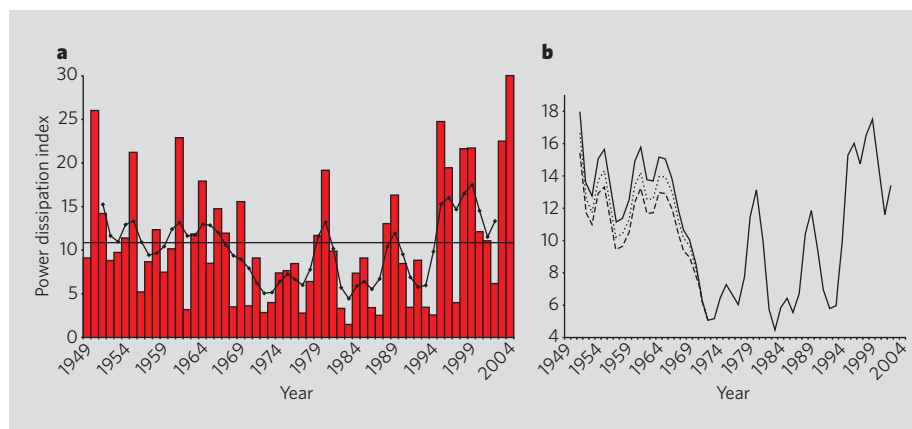


Figure 1 | Derivation of Atlantic power-dissipation index (PDI). **a**, Emanuel’s bias-correction version² of PDI for the North Atlantic tropical cyclones for 1949–2004. PDI takes into account frequency, duration and intensity of tropical cyclones by cubing the winds during the lifetime of the systems while they are of at least tropical-storm force (18 m s^{-1}) and summing them up for the year. Values shown are multiplied by 10^{-6} in units of $\text{m}^3\text{ s}^{-3}$. Horizontal line, time-series mean of 10.8; black curve, data after smoothing with two passes of a 1-2-1 filter. **b**, Three versions of the smoothed PDI for the North Atlantic using: dashed line, Emanuel’s applied bias-removal scheme; dotted line, 1993 version³ of the bias-removal scheme; solid line, original hurricane database. All three versions are identical from 1970 onwards.

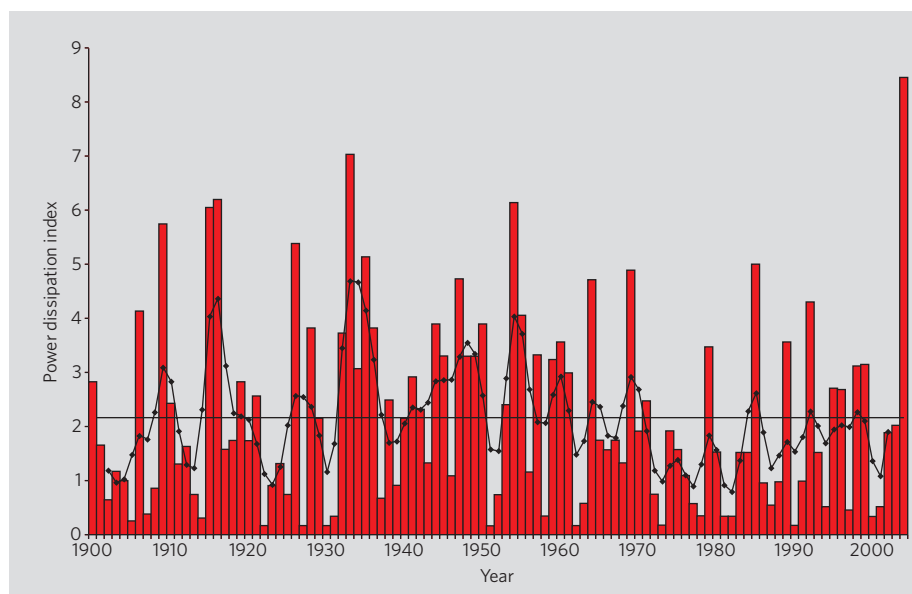


Figure 2 | The continental United States PDI at the time of impact for the reliable-period record of 1900–2004. This is computed from the best estimate of the peak sustained (1 min) surface (10 m) winds to have affected the US coastline for all tropical storms, subtropical storms and hurricanes causing at least gale-force (18 m s^{-1}) winds. Values shown are multiplied by 10^{-5} in units of $\text{m}^3\text{ s}^{-3}$. Horizontal line, time-series mean; black curve, data after smoothing with two passes of a 1-2-1 filter. For the continental US coast, the year 1900 roughly marks the start of a complete database. (Before that, portions of Florida, Louisiana and Texas were too sparsely settled to ensure adequate monitoring of all tropical cyclones, particularly those that were small but intense like 2004’s hurricane Charley.) The year 2004 stands out as the busiest from the twentieth century to the beginning of the twenty-first century, with 20% more PDI than the second most-active year in 1933. (However, 2004’s US PDI value is slightly less than that estimated to have occurred in 1886, as at least seven landfalling hurricanes struck that season, the busiest on record since 1851.)

1. Knutson, T. R. & Tuleya, R. E. *J. Clim.* **17**, 3477–3495 (2004).
2. Emanuel, K. *Nature* **436**, 686–688 (2005).
3. Landsea, C. W. *Mon. Weath. Rev.* **121**, 1703–1714 (1993).
4. Franklin, J. L., Black, M. L. & Valde, K. *Weath. Forecast.* **18**, 32–44 (2003).
5. Dunion, J. P., Landsea, C. W., Houston, S. H. & Powell, M. D. *Mon. Weath. Rev.* **131**, 1992–2011 (2003).
6. Kepert, J. & Wang, Y. *J. Atmos. Sci.* **58**, 2485–2501 (2001).

7. Kepert, J. J. *Atmos. Sci.* **58**, 2469–2484 (2001).
8. Landsea, C. W. et al. *Bull. Am. Meteorol. Soc.* **85**, 1699–1712 (2004).
9. Landsea, C. W., Pielke, R. A. Jr, Mestas-Núñez, A. M. & Knaff, J. A. *Clim. Change* **42**, 89–129 (1999).
10. Goldenberg, S. B., Landsea, C. W., Mestas-Núñez, A. M. & Gray, W. M. *Science* **293**, 474–479 (2001).
11. Landsea, C. W. et al. *Hurricanes and Typhoons: Past, Present and Future* (eds Murname, R. J. & Liu, K.-B.) 177–221 (Columbia Univ. Press, New York, 2004).
12. Blake, E. S., Rappaport, E. N., Jarrell, J. D. & Landsea, C. W. *The Deadliest, Costliest, and Most Intense United States Tropical Cyclones from 1851 to 2004 (and Other Frequently Requested Hurricane Facts)* (National and Oceanic Atmospheric Administration, Technical Memorandum NWS TPC-4, 2005).

doi:10.1038/nature04477

METEOROLOGY

Emanuel replies

Replying to: R. A. Pielke *Nature* **438**, doi:10.1038/nature04426 (2005) and C. W. Landsea *Nature* **438**, doi:10.1038/nature04477 (2005)

In my original Article¹, I showed that there has been a significant upward trend in a measure of tropical-cyclone power dissipation over the past 30 years¹. It is important to note that this measure is integrated over the life of the storm, and that the upward increase is evident in all major ocean basins prone to tropical cyclones. However, Pielke² finds no discernible trend in hurricane damage in the United States after correction for inflation and demographic trends, and Landsea³ finds no trend in US landfall-based hurricane power dissipation back to the turn of the last century.

Pielke suggests² that this apparent disparity could be explained if the power-dissipation trend I find is an artefact of the data and/or analysis methods, or if the trend is accurate but not a good predictor of damage. As this trend is large and universal — having about the same value in all the major ocean basins, despite different measurement techniques — and as it is well correlated with sea surface temperature (SST), which is relatively well measured, I stand by my conclusions about the trends in tropical-cyclone power dissipation.

I cannot discount the second of Pielke's conjectures, but the reason for the disparity may be more prosaic. Although Atlantic hurricanes do most of their destruction within 6–12 hours after landfall, they last for an average of 180 hours; moreover, only a fraction of hurricanes ever affect the US coastline. This means that the power-dissipation index (PDI) I used, which is accumulated over all storms and over their entire lives, contains about 100 times more data than an index related to wind speeds of hurricanes at landfall. There is large variability in wind speed over the life of each storm and large storm-to-storm random variability: detecting a temporal trend in the presence of this variability requires separation of the signal from the noise. With 100 times more data, my index has a signal-to-noise ratio that is ten times that of an index based on land-falling wind speeds. It is therefore possible that the real trend is detectable in the power dissipation but not in landfalling statistics. A simple calculation based on the observed root-mean-square variability of hurricane activity indicates that this is indeed the case,

and probably explains why Pielke² and Landsea³ find no trends in US landfall data.

Pielke argues that because El Niño can be detected in hurricane damage, a trend related to PDI should also be evident, if it exists. But the detectability of an El Niño signal in US hurricane damage is marginal, explaining only 3–4% of the variance⁴. Tropical Atlantic SST explains far more of the variance of both total Atlantic tropical-cyclone numbers and average tropical-cyclone intensity than does El Niño; but curiously, SST is even less correlated with a measure of US landfalling storm activity than El Niño. This probably once again reflects the difficulty of detecting trends in sparse time series in which the amplitude of random fluctuations is large compared with the signal.

The failure of any trend in landfall statistics to emerge from the noise is itself significant, and supports Pielke's view that demographic trends will be more important than climate change in coming years. But this is a short-term and US-centric view. When global tropical-cyclone activity is considered, and not just the 12% that occurs in the Atlantic region, a trend in landfalling intensity is already apparent; even in the Atlantic the signal, if it exists, is similar to the PDI trend, and if it continues should emerge from the noise in a few decades.

Landsea³ starts by saying that increasing SST has the potential for “slightly” increasing the intensity of tropical cyclones. But, as I discussed¹, the existing theory and modelling⁵ on which this assertion is based suggest that the predicted ~2 °C increase in tropical SST would increase wind speeds by 10% and, accounting for increased storm lifetime, increase power dissipation by 40–50%. This is hardly slight. The existing theory and modelling work⁵ are limited, however, in that they do not account for changes in environmental conditions, such as wind shear, and so only provide a loose guide as to what to expect.

Landsea correctly points out that in applying a smoothing to the time series, I neglected to drop the end-points of the series, so that these end-points remain unsmoothed. This has the effect of exaggerating the recent upswing in Atlantic activity. However, by

chance it had little effect on the western Pacific time series, which entails about three times as many events. As it happens, including the 2004 and 2005 Atlantic storms and correctly dropping the end-points restores much of the recent upswing evident in my original Fig. 1 and leaves the western Pacific series, correctly truncated to 2003, virtually unchanged. Moreover, this error has comparatively little effect on the high correlation between PDI and SST that I reported¹.

In correcting for biases in the original Atlantic tropical-cyclone data, I relied on a bias correction applied by Landsea⁶, presented as a table. I had fitted a polynomial to that correction, as I felt that a continuous rather than discrete correction was more defensible. Landsea believes that this had the effect of overcorrecting the most intense storms in the pre-1970 record, and I accept his revision to my analysis (Fig. 1b of ref. 3).

The Atlantic hurricane-intensity record by itself is not long enough to infer any connection between hurricanes and either global warming or multi-decadal cycles, but the high correlation between hurricane activity and tropical SST is remarkable (and largely unaffected by the corrections discussed), and the SST record is long enough to show the influence of global warming. To detect correlations with hurricane activity, tropical cyclones in the North Atlantic can be counted, assuming that detection of the presence of a storm by ships and islands is reliable (although intensity estimation is dubious before the mid-1940s). This count is highly correlated with both tropical Atlantic SST and Northern Hemispheric mean surface temperature through the entire record, casting doubt on whether the recent multi-decadal variability in tropical SST and hurricane activity is due purely to natural causes, as Landsea implies³.

I maintain that current levels of tropical storminess are unprecedented in the historical record and that a global-warming signal is now emerging in records of hurricane activity. This is especially evident when one looks at global activity and not just the 12% of storms that occur in the Atlantic. But I agree that there is a pressing need for a storm-by-storm reanalysis of tropical cyclones, not only in the North Atlantic, but also in the western North Pacific, where aircraft reconnaissance records also extend back to the 1940s.

Kerry Emanuel

Program in Atmospheres, Oceans, and Climate,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA
e-mail: emanuel@texmex.mit.edu

1. Emanuel, K. *Nature* **436**, 686–688 (2005).
2. Pielke, R. A. Jr *Nature* doi:10.1038/nature04426 (2005).
3. Landsea, C. W. *Nature* doi:10.1038/nature04477 (2005).
4. Katz, R. W. *J. Appl. Meteorol.* **41**, 754–762 (2002).
5. Knutson, T. R. & Tuleya, R. E. *J. Clim.* **17**, 3477–3495 (2004).
6. Landsea, C. W. *Mon. Weath. Rev.* **121**, 1703–1714 (1993).

doi:10.1038/nature04427

REVIEWS

An Asian perspective on early human dispersal from Africa

Robin Dennell¹ & Wil Roebroeks²

The past decade has seen the Pliocene and Pleistocene fossil hominin record enriched by the addition of at least ten new taxa, including the Early Pleistocene, small-brained hominins from Dmanisi, Georgia, and the diminutive Late Pleistocene *Homo floresiensis* from Flores, Indonesia. At the same time, Asia's earliest hominin presence has been extended up to 1.8 Myr ago, hundreds of thousands of years earlier than previously envisaged. Nevertheless, the preferred explanation for the first appearance of hominins outside Africa has remained virtually unchanged. We show here that it is time to develop alternatives to one of palaeoanthropology's most basic paradigms: 'Out of Africa 1'.

A key assumption in accounts of early hominin evolution is that the genus *Homo* originated in Africa, and an early form, classified either as *Homo ergaster* or *H. erectus sensu lato* (see Box 1), was the first to leave about 1.7–1.9 Myr ago (depending upon one's choice of dates and specimens), and then colonized southern Asia as far as 40° N. The identification of east Africa as the 'core' area for the genus *Homo* (including *H. ergaster*) as well as tool-making seems secure to most palaeoanthropologists, and the most recent attempts at modelling early hominin dispersals start implicitly from the assumption that *H. ergaster* originated in east Africa and then dispersed across Asia^{1,2}. In fact, the evidence that *H. ergaster* originated in east Africa is less convincing than it seems. *H. ergaster* marks such a radical departure from previous forms of *Homo* (such as *H. habilis*) in its height, reduced sexual dimorphism, long limbs and modern body proportions³ that it is hard at present to identify its immediate ancestry in east Africa⁴. Not for nothing has it been described as a hominin "without an ancestor; without a clear past"⁵.

Who were the first Asians?

At present, we have very little information on where, when and which hominins first appeared in Asia, and the expansion of *H. ergaster* across Asia in the Early Pleistocene remains a massive assumption, even if it is routinely treated as a historical fact. It is assumed that it migrated out of Africa along the Nile Valley or across the southern end of the Red Sea, but there is no archaeological or fossil hominin evidence that hominins were in the Nile Valley in the Lower Pleistocene; and there are no Oldowan sites in the Sinai, southern Negev, or in southwest Arabia at the alleged point of entry to Asia. The only Asian Early Pleistocene fossil hominin evidence comprises three incisors from 'Ubeidiya, Israel (1.4–1.0 Myr ago), attributed to *H. erectus sensu lato* (s.l.) by default⁶; the 1.7-Myr-old specimens from Dmanisi, Georgia⁷, which have recently been classified as a very early type of *H. ergaster*⁸ and/or a new taxon, *H. georgicus*⁹; and the specimens attributed to *H. erectus sensu stricto* (s.s.) in Java¹⁰, 5,300 miles away and regarded by many^{11–16} but not all^{10,17–20} as different from the east African *H. ergaster*. The earliest of these is the Mojokerto cranium²¹, which now seems to have been found in context, despite previous misgivings²², and is dated to 1.81 ± 0.04 Myr ago (ref. 23); the key specimens from Sangiran have been dated to about

1.6–1.7 Myr ago (ref. 23, 24). (The Early Pleistocene mandible and teeth attributed to *Homo* from Longgupo²⁵, southern China, probably belonged to an ape^{26,27}.) This meagre list of sites is supplemented by archaeological instances of Early Pleistocene artefacts in Asia that are attributed to *H. erectus s.l.*; examples are Erq el-Ahmar, Israel, claimed to date to the Olduvai Event²⁸, and the Nihewan basin, north China²⁹. The only reason why the earliest tool assemblages in Asia are attributed to *H. erectus s.l.* is that palaeoanthropologists have already decided that, in effect, it was the only hominin capable of migration out of Africa, and with sufficient *Wanderlust*¹³ to do so.

Box 1 | *Homo erectus* and *Homo ergaster*

H. erectus—or more properly, *Pithecanthropus erectus*—was first discovered at Trinil in Java by Eugène Dubois in 1891. In the 1930s, further discoveries of hominin remains elsewhere in Indonesia and at Chou-kou-tien (now Zhoukoudian) in China were seen as broadly similar, even if initially given their own generic names (such as *Meganthropus* and *Sinanthropus*). In 1950, Ernst Mayr reclassified all this material as *H. erectus*, with the Trinil specimens as the type fossils. Subsequent African specimens were also called *H. erectus*, as were much later specimens from Europe. *H. erectus* thus became for a while the earliest hominin that was thought to have lived in Asia, Africa and Europe.

In the past few years, some have doubted that the east African specimens should be placed within the same palaeospecies as the Asian ones. In the light of discoveries at Koobi Fora, it has been suggested that the earliest African examples should be called *H. ergaster*, after the specimens found at Koobi Fora, including WT15000, the magnificent 1.6-Myr-old skeleton of a young boy from Nariokotome that was initially published as *H. erectus*. Consequently, it is the African *H. ergaster* that is now seen by some as the hominin that first colonized Asia and formed the founding population of what later became *H. erectus* in China and southeast Asia. European specimens once regarded as late examples of *H. erectus* or 'archaic *Homo sapiens*' are now increasingly classified under the revived taxon of *H. heidelbergensis*, a term first used to classify the mandible from Mauer, Germany, found in 1907. To avoid ambiguity, the term *H. erectus* is used here *sensu stricto* to denote only those specimens from eastern Asia, and *H. ergaster* is used to denote early east African specimens of *H. erectus sensu lato*.

¹Department of Archaeology, University of Sheffield, Sheffield S1 4ET, UK. ²Department of Archaeology, Leiden University, PO Box 9515, 2300RA Leiden, The Netherlands.

Homo ergaster: wanderer or stay-at-home?

The reason why *H. ergaster* is assumed to have been uniquely capable of migrating out of Africa about 1.7–1.9 Myr ago into the Asian grasslands is because of its long limbs, human-like body proportions, probable efficient thermoregulatory mechanisms for remaining cool in hot conditions, the ability to ingest large amounts of meat in an environment rich in fauna but poor in plant foods for a hungry primate, and a sufficiently large brain to deal with the challenges of a more carnivorous niche³. This argument is persuasive, except for the point that australopithecines had probably colonized all the African savannah grasslands by 3.0–3.5 Myr ago (ref. 30), and *Australopithecus garhi* was living in a similar environment in north-east Africa by 2.5 Myr ago (ref. 31). As savannah grasslands were extensive across southern Asia by 3.0 Myr ago (ref. 32), there are no reasons *a priori* why australopithecines could not also have expanded into the Asian grasslands before *H. ergaster*.

The fossil evidence that *H. ergaster* was in Asia in the Early Pleistocene is not only weak, but extremely ambiguous. The long-running debate (see Box 1) over whether the Early Pleistocene Javan hominins should be classified as *H. erectus* s.s. (and thus different from their east African counterparts) or incorporated with them as *H. erectus* s.l., and/or seen as composite³³ is still unresolved. Neither those who regard them as an integral part of *H. erectus* s.l.¹⁰ nor those who view them as derived from the African *H. ergaster* question that the core area was east Africa. Neither position is strengthened by the recent suggestion that the Mojokerto child had an ape-like pattern of postnatal development³⁴, unlike *H. ergaster*. Another ‘big unknown’ is the ancestral form of the Late Pleistocene *H. floresiensis*^{35,36}, which may prove to have roots deep in the Asian Pleistocene. The Dmanisi hominins are harder still to assimilate within a simple model of African origin and Asian dispersal by *H. ergaster* or *H. erectus* s.l. The first discovery, of the mandible, could be classified as *H. erectus* s.l.³⁷. The first crania were regarded as a very early form of *H. ergaster*⁸, and the latest, very small-brained one (D2700) as *H. ergaster* but also closely related to *H. habilis* s.s. (ref. 38). Confusingly, one mandible (D2600) has been assigned to a new taxon, *H. georgicus*⁹. These recent assessments imply that hominins dispersed from Africa earlier than the emergence of large-bodied hominins such as the Nariokotome individual.

When could hominins first have left Africa?

But how much earlier could this dispersal have been? If a hominin at the same grade as *H. habilis* was able to exist outside Africa, why not others? Why not follow the logic of Wood and Collard’s reasoning³⁹, that *H. habilis* is better classified as *A. habilis* and suggest that the earliest Asians were in fact australopithecine, with *A. georgicus* as their first known representative outside Africa? This suggestion would open a Pandora’s box: if a hominin as small-brained (and probably as short) as those at Dmanisi could colonize southwest Asia by 1.7 Myr ago (and with no obvious African antecedents), why not at 2.6 Myr ago, shortly after stone tool making became part of the hominin repertoire⁴⁰? Or why not even earlier, by, for example, 3.0–3.5 Myr ago, when the Saharan–Arabian desert barriers did not yet exist³²? If *A. bahrelghazali*, 2,500 km west of the Rift Valley, implies that by 3.5 Myr ago “hominids were distributed throughout the woodland and savannah belt from the Atlantic Ocean across the Sahel through eastern Africa to the Cape of Good Hope”³⁰, why could they not have done the same across the grasslands of western, southern and central Asia?

Absence of evidence and evidence of absence

The obvious retort to these questions is that there is no evidence that australopithecines did migrate out of Africa. However, absence of evidence is not enough; we need convincing evidence (so far not forthcoming) that the absence is not the result of taphonomic circumstance or lack of fieldwork, especially in a continent as large as Asia. There are only a limited number of vertebrate fossil

assemblages for the Late Pliocene and Early Pleistocene of southwest Asia (a region larger than Kenya, Ethiopia and Tanzania combined). The Late Pliocene assemblage from Bethlehem⁴¹ (Israel) is very small, with only 11 taxa and dominated by animals with an adult body weight of more than 60 kg and thus larger than hominins. These fossils were found in coarse gravel (with clasts up to 0.5 m long) in a clay matrix, in which small and fragile remains were most unlikely to be preserved. At Kvabebi^{42,43}, Georgia, dated to more than 2.6 Myr ago (that is, earlier than the earliest stone tools in Africa⁴⁰), there are 21 mammalian taxa indicative of a riverine and marshy environment. Two other Georgian localities, Kočachuri and Calka, are slightly earlier than Dmanisi and yielded small assemblages dominated by large taxa⁴². There are 21 taxa represented at Dmanisi, and 33 (excluding microfauna) at ‘Ubeidiya. The point here is that small assemblages, with only a few taxa that are mainly from large animals, are most unlikely to contain hominin remains: in southwest Asia, two of the three large and fine-grained assemblages did yield skeletal evidence of hominins (although very little at ‘Ubeidiya⁶). In central Asia, the Late Pliocene record is poor because fossils are often found in coarse sediments⁴⁴, although one surprising find is of the baboon (*Papio suschkini*)⁴⁵ that is often regarded as a commensal of *Homo*⁴⁶. Late Pliocene faunal assemblages from northern Pakistan and India do not contain any hominins, and these are also unknown for the entire Early Pleistocene: the earliest fossil hominin evidence we have is Middle Pleistocene, from the Narmada Valley⁴⁷, long after hominins are first in evidence to both the west and east of the Indian subcontinent. Although many vertebrate taxa are represented in the Upper Siwaliks of India (30 in the Tatrot and 49 in the succeeding Pinjor Stage⁴⁸), most are larger than hominins. It is also likely that the full range of taxa is incomplete for the Indian subcontinent, because *Megantereon* and *Pachycrocuta* are not recorded in India but are present in Pakistan; in Pakistan, there is no evidence of *Camelus* and small primates, and in neither country is *Homotherium* recorded⁴⁹, although this is present to the west at Dmanisi, to the north at Kuruksay, central Asia⁴⁴ and to the east at Longgupo²⁵, south China. In mainland southeast Asia, there is no Late Pliocene or Early Pleistocene fossil evidence. One of the few instances for which we can be reasonably sure that *H. erectus* s.l. (and other hominins) were absent is Longgupo²⁵, south China, where four primates (*Gigantopithecus*, *Lufengpithecus*, *Macaca* and *Procyoncocephalus*) are recorded among the 68 taxa present. The faunal assemblages from the Yushe Basin⁵⁰ and the “Hipparion fauna” of north China⁵⁰ are of comparable quality to those from India, and the absence of hominins is equivocal.

Ways forward: alternatives to Out of Africa 1

If the above taphonomic review suggests that we cannot show the absence of hominins from areas in Asia at a time before the little evidence we have indicates their presence, we need to consider alternatives to the current Out of Africa model. There are three issues here. The first is when hominin(s) first left Africa—might they, for example, have left shortly after they acquired the ability to make stone tools, the earliest of which are currently 2.6 Myr old? Or could they have left even earlier, about 3.0–3.5 Myr ago, when some australopithecines were already living in the African grasslands? The second issue is whether we yet know the full range of hominins that inhabited both Africa and Asia in the Late Pliocene and Early Pleistocene. Even in east Africa, several new taxa have been claimed in the past decade (for example, *A. anamensis*⁵¹, *A. garhi*⁵², *Ardipithecus ramidus*⁵³ and *Kenyanthropus platyops*⁵⁴) and doubtless more will be found. (An indication of how little we know about Pleistocene east Africa is that only recently has the first fossil evidence for chimpanzee been found⁵⁵.) In Asia, the recent discoveries of *H. georgicus* and *H. floresiensis* should make us very wary of assuming that *H. erectus* s.l. was the only player on the Asian stage in the Early Pleistocene. Third, Asia might not have been the passive recipient of whatever migrated out of Africa but might have been a major donor to

speciation events, as well as dispersals back into Africa. Such two-way traffic is well documented for other mammals in the Pliocene and Early Pleistocene, such as *Equus*⁵⁶ and bovids⁵⁷, with more taxa migrating into than out of Africa. There is no reason why hominin migrations were always from Africa into Asia, and movements in the opposite direction might also have occurred, as has been suggested for the Olduvai OH9 (refs 13, 58) and Daka⁴ specimens. We should even allow for the possibility that *H. ergaster* originated in Asia^{23,59} and perhaps explain its lack of an obvious east African ancestry as the result of immigration rather than a short (and undocumented) process of anagenetic (*in situ*) evolution. Although Darwin's suggestion⁶⁰ that "it is somewhat more probable that our early progenitors lived on the African continent than elsewhere" is widely quoted, it is worth noting his next sentence: "But it is useless to speculate on this subject ... since so remote a period the earth has certainly undergone many great revolutions, and there has been ample time for migration on the largest scale".

We obviously need ways of testing these alternatives. The overriding need is for data sets from Asia that are of comparable quality to those from Africa. Although absence can never be 'demonstrated', we can at least put some constraints on its probability by comparing the quality of the fossil record from the inferred core and its 'peripheries'. Because the species we are interested in were not very abundant in their world and hence in the fossil record, knowledge of biasing factors is a prerequisite for the study of their past distributions. The higher the trophic level of a species, the smaller is its abundance in the real world and hence in the fossil record (for the whole of the Asian Late Pliocene, we have only two records of a puma, for example, separated by more than 3,000 miles)⁴³. Open, mesic-to-arid environments tend to preserve fossils better than do forested and wetter environments (which is probably why we have no fossil record for the gorilla and only one observation—from open woodland—for the chimpanzee⁵⁵). Should faunal remains get covered in a sedimentary matrix, that matrix obviously has to survive and be accessible, a condition that is rarely met for Pliocene and Pleistocene sediments. The Rift Valley constitutes a unique exception by its sheer size and the exposure of fine-grained sediments of the relevant age, and includes many of the key African sites, as well as Erq el-Ahmar, 'Ubeidiya and Gesher Benot Ya'aqov in Israel.

Regional imbalances and future challenges

Current large-scale imbalances between regional records are often the result of differences in research history and intensity. To some degree the Javan and Levantine records result from research initiated during colonial times, and the east African record similarly owes a great deal of its incipient (and prolific) research to the consequences of its colonial history. In contrast, most parts of Asia have experienced only a very limited survey of Neogene exposures, in comparison with the heavy palaeoanthropological investments made in east Africa during the past four decades. These regional imbalances are crucial. After all, in the early twentieth century, east Asia was thought to have been the centre of human origins because it had the oldest fossils, and one of the marginal areas, Africa, had not yet seen any significant field-work⁶¹. Despite the imbalance in research intensity (and hence number of sites), Asia has produced major surprises in recent years, a testimony to its palaeoanthropological potential. As an example, recent research in China has extended its earliest hominin presence up to 1.66 Myr ago (ref. 29), half a million years older than envisaged only a few years ago, with the lowest levels of the fossiliferous sequence in the Nihewan basin not having been reached yet. The increasing evidence for Early Pleistocene hominins in China and Java stretches the limits of current thinking on hominin evolution, as do the finds of *H. floresiensis* and the Dmanisi hominin assemblage, the latter recovered from an area where few would have expected Early Pleistocene tiny-brained hominins two decades ago. These discoveries underscore our poor ability to discern, let alone predict, the design on the picture we try to piece together from the

few pieces of the jigsaw that we have. Again, absence of evidence is not enough: if we postulate that species A migrated into area B, we need comparable data sets to infer legitimately that it was absent before that date—we need not just the FAD (first appearance date) of a taxon in a new area, but also its LPA (last probable absence) (see Fig. 1). We will never have certainty about LPAs, of course: the recently reported finds from Pakefield, southern England, show that two centuries of intensive research of the Cromer Forest beds failed to recover the (indeed ephemeral) traces of an early Middle Pleistocene hominin presence there⁶². That such a surprise can turn up in one of the best-researched areas of the Old World shows that we can never be sure about LPAs and, more importantly, underlines the necessity of working with data sets of comparable quality in both alleged donor and alleged recipient regions (see also Box 2). For Africa and Asia, comparability is still many generations of research grants away. Nevertheless, we could do much more to reduce the level of uncertainty over when hominins were last absent in Asia by increasing the number and quality of fossil assemblages immediately before their first alleged appearance.

'Africa' and 'Asia', or 'Savannahstan'?

We also need different spatial units for investigating extinct hominin populations. Since the time of the Greeks and the Romans, we habitually refer to 'Africa' and 'Asia' as separate continents, each somehow homogeneous and distinct from the other. Plants and animals (and extinct hominins) are less respectful of our Graeco-Roman heritage. The landmasses we now call Africa and Asia are of course enormously diverse, but they also have many plants, animals

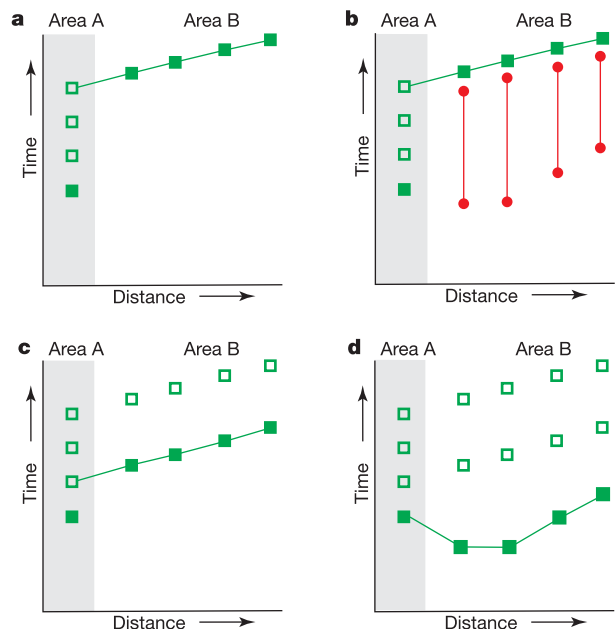


Figure 1 | Dispersals, cores and peripheries. The dangers of over-reliance on first appearance dates (FADs) of when a taxon migrated from its core area, A, into a new territory, B. Filled green squares indicate the first appearance of a taxon, open green squares the presence of a taxon and red circles fossil assemblages without this taxon. **a**, A hypothetical situation in which a taxon originated in area A, and then migrated into territory B. **b**, The reliability of these FADs is considerably strengthened by the numerous well-dated instances when their last probable absences (LPAs) can be documented. Without these, future discoveries might indicate (as shown in **c**) that previous estimates of when a taxon first appeared were too recent, as happened when the earliest Javan hominins were redated from about 1.0 Myr old to 1.8 Myr old. **d**, Even more alarmingly, future discoveries might even show that the taxon probably originated in the area that it was supposed to have colonized—as happened when the centre of hominin origins was relocated from Asia to Africa in the 1960s.

and environments in common. In recent decades, palaeoanthropologists have rightly emphasized the importance of savannah grasslands in hominin evolution, both as a place where many types (including *H. ergaster*) lived in the Late Pliocene and Early Pleistocene and as having been important in influencing hominin brain size, post-cranial anatomy, and diet. As noted earlier, Pliocene grasslands extended all the way from west Africa to north China, and 'Savannahstan' might prove a more useful spatial unit for modelling early hominin adaptations and dispersals within them than simply an undifferentiated 'Africa' or 'Asia'. For example, the African hominins 1.9–1.7 Myr ago at Koobi Fora (Kenya) and Ain Hanech (Algeria), and their slightly later counterparts in Asia at 'Ubeidiya (Israel), and Majuangou (north China) were all living in broadly comparable grassland environments, and it makes sense to place them within the same frame of reference. This might also highlight significant

Box 2 | Neanderthals and moderns—where were they?

Highlighting the imperfections of the fossil record obviously has implications for studies of the distribution of other species too. The Neanderthals are by far the best-studied extinct hominins, with a rich fossil record sampling hundreds of individuals, heavily biased towards the western part of their range, western Europe. However, the northern, eastern and southern limits to their distribution are poorly documented, again because of an imbalance in research intensity. The juvenile from Teshik-Task, Uzbekistan, is the easternmost one known, at roughly 1,300 and 2,000 miles from its nearest fossil neighbours, Shanidar in Iraq and Kiik-Koba in the Black Sea area, respectively. The southern limit of their distribution is unknown, and may have extended over the whole of Arabia and the Indian subcontinent—we cannot be certain until these regions produce the necessary fossil evidence. Distribution maps of Neanderthals are palimpsests of range expansion and contraction, probably hide many shifts of ranges in the rhythm of climatic oscillations and most probably also understate their full range. Chronological resolution is significantly better than in the earlier Pleistocene, however, and enables us to see the occasional 'interfingering' of their range with that of anatomically modern humans in the Near East. If the Tabun C1 hominin does indeed date to Marine Isotope Stage (MIS) 6, Neanderthals were in the northern parts of the Near East before anatomically modern humans were there in MIS 5, with Neanderthals again present during MIS 4. As with Out of Africa 1, the poor state of sampling of major parts of western and central Asia should force us to be very humble in our inferences on the core and peripheries in the Neanderthal world.

Comparable points can be made with regard to Out of Africa 2 (the model according to which our own species, *H. sapiens* (or 'anatomically modern humans') evolved in Africa by 150–200 kyr ago and then migrated outwards into Asia and Europe and eventually replaced all indigenous populations in these regions). Current evidence indicates that anatomically modern humans appeared in east Africa by about 200 kyr ago (the new dates for Omo-Kibish), whereas the earliest outside Africa are those from Israel at about 115 kyr ago. There are no fossil hominins of the same age as Omo-Kibish from southwest, south or central Asia, so it is an open question as to when they first appeared in those regions: possibly at 125 kyr ago, perhaps 200 kyr ago, or even earlier. (The only potential evidence is the specimen from Zuttiyeh Cave, Israel, which is not particularly diagnostic and is almost certainly considerably older than the associated Th-U date of 154 kyr ago.) All we know about southwest Asian hominins between 300 and 125 kyr ago is that Neanderthals were present during part of MIS 6. As with Out of Africa 1, if we cannot show when modern humans were last absent in a region, we have no secure means of knowing from the skeletal record whether that region was core or peripheral. Although genetic studies of modern humans strongly imply that they originated in Africa, these studies are notoriously vague as to when they first dispersed, or even whether the modern genotype dispersed without any population expansion from Africa. Current fossil evidence from Asia is clearly incapable of testing these suggestions at present.

variation that is sometimes buried under a blanket term such as 'Asia'. For example, the hominins in the Nihewan basin, north China, and those in Java are both clearly in east Asia, but those in Java inhabited a region that was considerably more densely wooded. It is not the continent that matters in studying human origins so much as the type(s) of environment with which early hominins were associated.

Hominins, not just *Homo*, outside Africa

We also need to focus more on hominins and not just *Homo* when studying early hominins outside Africa. Archaeological approaches to early lithic assemblages in Asia are a good case here. Any stone tool assemblage in Asia dated as older than 1.9 Myr ago (the earliest date that *Homo* is supposed to have left Africa) is either dismissed or (more usually) ignored⁶³; undated Oldowan tools are assumed to date from after 1.9 Myr ago and not from 2.6 Myr ago (the date of their first appearance in east Africa); and stone tool assemblages in Asia dated to the Olduvai Event (1.77–1.95 Myr ago) and not associated with hominin remains are automatically attributed to *Homo erectus* s.l. However, there is no reason why Oldowan assemblages in Arabia cannot be older than 1.9 Myr old, or why the tools from Ain Hanech⁶⁴ (Algeria) or Erq el Ahmar (Israel) were made by *H. erectus* s.l., not least because similar assemblages were made in east Africa at that time (and earlier in some cases) by *H. habilis*, *H. ergaster* and probably *H. rudolfensis*, *A. garhi* and *Paranthropus*. We may be due for some big surprises in discovering that *H. ergaster* was not the only, or even the first, African tool-making hominin to leave home.

Human evolution writ new?

As readers of this journal will be aware, Asia has produced some surprising discoveries in the past decade, including two new palaeo-species of *Homo*. Recent African discoveries in Chad are also highly pertinent to those interested in early Asian hominins. The discovery of *Sahelanthropus tchadensis*⁶⁵ shows that hominins were already well beyond the east African Rift in the late Miocene, and *A. bahrelghazali* indicates that the African grasslands were probably colonized by 3.0–3.5 Myr ago. The latter discovery raises obvious implications for when the Asian grasslands were first colonized, and whether large

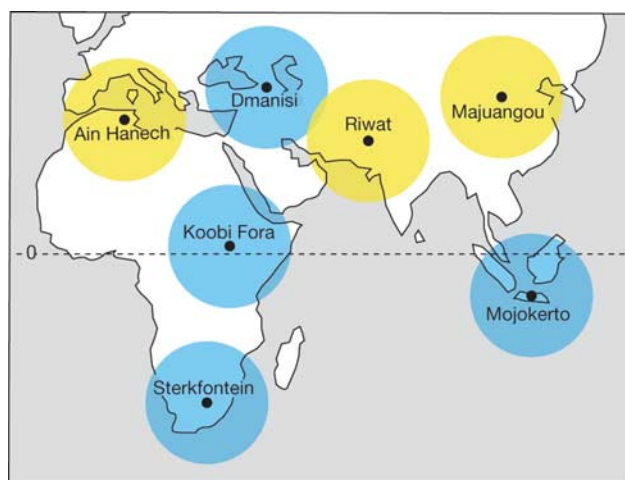


Figure 2 | The hominin world about 1.7 Myr ago. The circles denote radii of 1,000 miles. Blue circles indicate known populations of *Homo* at this time: *H. ergaster* and various australopithecines (including *A. (H.) habilis* in east Africa, *H. georgicus* in Georgia, *H. erectus* in Java, and *Paranthropus* and *Homo* in southern Africa). The ancestries of *H. ergaster*, *H. georgicus* and *H. erectus* are unclear, as are their spatial extents and relationships to each other. The yellow circles indicate areas with no fossil hominin evidence but where stone tools were being made: north China, Algeria and Pakistan. As shown, there is ample 'ecological space' in Asia for more hominins than currently recorded.

brains, modern body size and proportions, and obligate bipedalism were essential for that process to occur. These recent finds are not easily reconciled with the notion that hominins originated in the Rift Valley and that *H. ergaster* was the first and only hominin to migrate out of Africa in the Late Pliocene/Early Pleistocene. Most probably, we are on the threshold of a profound transformation of our understanding of early hominin evolution that might prove as far-reaching as the demise of the notion of *Man the Hunter*⁶⁶ in the early 1960s. The process was often painful and accompanied by heated debate, but our understanding of early hominin subsistence improved enormously (and, indeed, some parts of the original model were strengthened by it⁶⁷). Although there will doubtless be an understandable reluctance to abandon Out of Africa 1, in its present form, as a model that is widely accepted as adequate for explaining a very small amount of data from Asia, there are benefits to be gained by widening our range of possible hypotheses. The present model is stifling a rigorous evaluation of how we can interpret the sparse, but recently much improved, data upon which it rests. Useful initial steps would be for us to be more explicit about just how few reliable observations we have of Early Pleistocene hominins across the Old World. Other steps would be to pay more attention to the comparability of data sets when evaluating whether or not the absence of hominins is more than the outcome of taphonomic circumstance or the history of fieldwork; additionally, an emphasis on different spatial units and on hominins other than *H. ergaster* (including earlier ones) might prove fruitful. Meanwhile, if we cannot demonstrate the probable absence of a hominin (including *H. erectus*) in a region, we should reserve judgement as to when it first appeared there. Another useful step would be to dispense with most of the arrows indicating movement from alleged (but often unproven) core territories into (alleged but often unproven) peripheral ones. It might be more profitable instead to focus on populations⁶⁸ as the basic unit of study (see Fig. 2), each of which might have its own local origins and history, and to accept that the boundaries of each, and the relations between them, are at present unknown, because all we have are isolated sampling points. Although these changes would radically alter the way in which we view human evolution outside Africa, they might prove more fruitful than continuing to envisage the earliest evidence for hominins in Asia as the outcome of a conjectural migration, from an unproven centre of origin and along hypothetical routes of dispersal.

- Antón, S. C., Leonard, W. R. & Robertson, M. L. An ecomorphological model of the initial hominid dispersal from Africa. *J. Hum. Evol.* **43**, 773–785 (2002).
- Mithen, S. & Reed, M. Stepping out: a computer simulation of hominid dispersal from Africa. *J. Hum. Evol.* **43**, 433–462 (2002).
- Aiello, L. C. & Wells, J. C. K. Energetics and the evolution of the genus *Homo*. *Annu. Rev. Anthropol.* **31**, 323–338 (2002).
- Asfaw, B. *et al.* Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature* **416**, 317–320 (2002).
- Walker, A. & Shipman, P. *The Wisdom of Bones: In Search of Human Origins* (Weidenfeld & Nicholson, London, 1996).
- Belmaker, M., Tchernov, E., Condemi, S. & Bar-Yosef, O. New evidence for hominid presence in the Lower Pleistocene of the Southern Levant. *J. Hum. Evol.* **43**, 43–56 (2002).
- Lordkipanidze, D. *et al.* The earliest toothless hominin skull. *Nature* **434**, 717–718 (2005).
- Gabunia, L. *et al.* Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* **288**, 1019–1025 (2000).
- Gabunia, L., de Lumley, M.-A., Vekua, A., Lordkipanidze, D. D. & de Lumley, H. Découverte d'un nouvel hominidé à Dmanisi (Transcaucasie, Géorgie). *C. R. Palévol.* **1**, 243–253 (2002).
- Antón, S. C. Natural history of *Homo erectus*. *Am. J. Phys. Anthropol.* **122**, 126–170 (2003).
- Andrews, P. J. An alternative interpretation of the characters used to define *Homo erectus*. *Courier Forschunginst. Senckenb.* **69**, 167–175 (1984).
- Foley, R. & Lahr, M. Human evolution writ small. *Nature* **431**, 1043–1044 (2004).
- Tattersall, I. Out of Africa again... and again? *Sci. Am.* **276**, 46–53 (1997).
- Tattersall, I. & Schwartz, J. *Extinct Humans* (Nevraumont, New York, 2000).
- Wood, B. & Turner, A. Out of Africa and into Asia. *Nature* **378**, 239–240 (1995).
- Wood, B. & Collard, M. The changing face of genus *Homo*. *Evol. Anthropol.* **8**, 195–207 (1999).
- Bilborough, A. in *Structure and Contingency: Evolutionary Processes in Life and Human Society* (ed. Bintliff, J.) 43–101 (Leicester Univ. Press, London, 1999).
- Brauer, G. How different are Asian and African *Homo erectus*? *Courier Forschunginst. Senckenb.* **171**, 301–318 (1994).
- Kramer, A. Human taxonomic diversity in the Pleistocene: Does *Homo erectus* represent multiple hominid species? *Am. J. Phys. Anthropol.* **91**, 161–171 (1993).
- Rightmire, G. P. *The Evolution of Homo erectus: Comparative Anatomical Studies of an Extinct Human Species* (Cambridge Univ. Press, Cambridge, 1990).
- Antón, S. C. Developmental age and taxonomic affinity of the Mojokerto child, Java, Indonesia. *Am. J. Phys. Anthropol.* **102**, 497–514 (1997).
- Huffman, F., Shipman, P., Hertler, C., de Vos, J. & Aziz, F. Historical evidence of the 1936 Mojokerto skull discovery, East Java. *J. Hum. Evol.* **48**, 321–363 (2005).
- Swisher, C. C. *et al.* Age of the earliest known hominids in Java, Indonesia. *Science* **263**, 1118–1121 (1994).
- Larick, R. *et al.* Early Pleistocene 40Ar/39Ar ages for Bapang Formation hominins, Central Java, Indonesia. *Proc. Natl Acad. Sci. USA* **98**, 4866–4871 (2001).
- Wanpo, H. *et al.* Early *Homo* and associated artefacts from Asia. *Nature* **378**, 275–278 (1995).
- Schwartz, J. H. & Tattersall, I. Whose teeth? *Nature* **381**, 201–202 (1996).
- Wu, X. Longgupo mandible belongs to ape. *Acta Anthropol. Sin.* **19**, 1–10 (2000).
- Ron, H. & Levi, S. When did hominids first leave Africa? New high-resolution paleomagnetic evidence from the Erk-El-Ahmar formation, Israel. *Geology* **29**, 887–890 (2001).
- Zhu, R. X. *et al.* New evidence on the earliest human presence at high northern latitudes in northeast Asia. *Nature* **431**, 559–562 (2004).
- Brunet, M. *et al.* The first australopithecine 2,500 kilometres west from the Rift Valley (Chad). *Nature* **378**, 273–275 (1995).
- De Heinzelin, J. *et al.* Environment and behaviour of 2.5-million-year-old Bouri hominids. *Science* **284**, 625–629 (1999).
- Dowsett, H. *et al.* Joint investigations of the Middle Pliocene climate. I: PRISM palaeoenvironmental reconstructions. *Global Planet. Change* **9**, 169–195 (1994).
- Tyler, D. E. An examination of the taxonomic status of the fragmentary mandible Sangiran 5, (*Pithecanthropus dubius*), *Homo erectus*, 'Meganthropus', or 'Pongo'? *Quat. Int.* **117**, 125–130 (2004).
- Coqueugniot, H., Hublin, J.-J., Veillon, F., Houët, F. & Jacob, T. Early brain growth in *Homo erectus* and implications for cognitive abilities. *Nature* **431**, 299–302 (2004).
- Brown, P. *et al.* A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
- Morwood, M. J. *et al.* Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* **437**, 1012–1017 (2005).
- Gabunia, L. & Vekua, A. A Plio-Pleistocene hominid from Dmanisi, East Georgia, Caucasus. *Nature* **373**, 509–512 (1995).
- Vekua, A. *et al.* A new skull from Dmanisi, Georgia. *Science* **297**, 85–89 (2002).
- Wood, B. & Collard, M. The human genus. *Science* **284**, 65–71 (1999).
- Semaw, S. *et al.* 2.6-Million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. *J. Hum. Evol.* **45**, 169–177 (2003).
- Hooijer, D. A. An early Pleistocene mammalian fauna from Bethlehem. *Bull. Br. Mus. Nat. Hist. Geology* **3**, 267–292 (1958).
- Vekua, A. Die Wirbeltierfauna des Villafranchium von Dmanisi und ihre biostratigraphische Bedeutung. *Jb. Röm.-Ger. ZentMus. Mainz* **42**, 77–180 (1995).
- Hemmer, H., Kahlke, R.-D. & Vekua, A. K. The Old World puma—*Puma pardoides* (Owen, 1846) (Carnivora: Felidae)—in the Lower Villafranchian (Upper Pliocene) of Kvabebi (East Georgia, Transcaucasia) and its evolutionary and biogeographical significance. *N. Jb. Geol. Paläont. Abh.* **233**, 197–231 (2004).
- Sotnikova, M. V., Dodonov, A. E. & Pen'kov, A. V. Upper Cenozoic bio-magnetic stratigraphy of Central Asian mammalian localities. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **133**, 243–258 (1997).
- Maschenko, E. N. *Papio (Paradolichopithecus) suschkini*: a revision of systematics, morphofunctional peculiarities of the skull and mandible [in Russian]. *Paleotierologiya*, 15–57 (1994).
- Bohe, R. & Behrensmeyer, A. K. The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **207**, 399–420 (2004).
- Sonkalia, A. Skull cap of an early man from the Narmada Valley alluvium (Pleistocene) of Central India. *Am. Anthropol.* **87**, 612–616 (1995).
- Nanda, A. C. Upper Siwalik mammalian faunas of India and associated events. *J. Asian Earth Sci.* **21**, 47–58 (2002).
- Dennell, R. *Early Hominin Landscapes in Northern Pakistan* (BAR/Archaeopress, Oxford, 2004).
- Flynn, L. J., Tedford, R. H. & Zhanxiang, Q. Enrichment and stability in the Pliocene mammalian fauna of North China. *Paleobiology* **17**, 246–265 (1991).
- Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376**, 565–571 (1995).

52. Asfaw, B. *et al.* *Australopithecus garhi*: A new species of early hominid from Ethiopia. *Science* **284**, 629–635 (1999).
53. White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
54. Leakey, M. G. *et al.* New hominin genus from eastern Africa shows diverse Middle Pleistocene lineages. *Nature* **410**, 433–439 (2001).
55. McBrearty, S. & Jablonski, N. G. First fossil chimpanzee. *Nature* **437**, 105–108 (2005).
56. Lindsay, E. H., Opdyke, N. D. & Johnson, N. M. Pliocene dispersal of the horse *Equus* and late Cenozoic mammalian dispersal events. *Nature* **287**, 135–138 (1980).
57. Vrba, E. S. in *Paleoclimate and Evolution* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 385–424 (Yale Univ. Press, New Haven, 1995).
58. Clarke, R. J. Out of Africa and back again. *Int. J. Anthropol.* **15**, 185–189 (2000).
59. White, T. D. in *Paleoclimate and Evolution* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 369–385 (Yale Univ. Press, New Haven, 1995).
60. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
61. Dennell, R. W. in *Studying Human Origins. Disciplinary History and Epistemology* (eds Corbey, R. & Roebroeks, W.) 45–66 (Amsterdam Univ. Press, Amsterdam, 2001).
62. Parfitt, S. A. *et al.* The earliest record of human activity in northern Europe. *Nature* **438**, 1008–1012 (2005).
63. Dennell, R. W., Rendell, H. M. & Hailwood, E. Early tool-making in Asia: two-million year-old artefacts in Pakistan. *Antiquity* **62**, 98–106 (1988).
64. Sahnouni, M. *et al.* Further research at the Oldowan site of Ain Hanech, North-eastern Algeria. *J. Hum. Evol.* **43**, 925–937 (2002).
65. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
66. Lee, R. B. & DeVore, I. (eds) *Man the Hunter* (Aldine Publishing Company, Chicago, 1968).
67. Dominguez-Rodrigo, M. & Pickering, T. R. Early hominid hunting and scavenging: a zooarcheological review. *Evol. Anthropol.* **12**, 275–282 (2003).
68. Howell, F. C. in *Contemporary Issues in Human Evolution* (eds Meikle, W. E., Howell, F. C. & Jablonski, N. G.) 1–45 (California Acad. Sciences, San Francisco, 1996).

Acknowledgements We thank various colleagues for comments on earlier drafts of this paper. R.D. thanks the British Academy for a three-year research professorship for his research into Asian prehistory. This work was supported by an internationalization grant of the Netherlands Organisation for Scientific Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to R.D. (r.dennell@sheffield.ac.uk).

Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*

James E. Galagan¹, Sarah E. Calvo¹, Christina Cuomo¹, Li-Jun Ma¹, Jennifer R. Wortman², Serafim Batzoglou³, Su-In Lee³, Meray Bastürkmen⁴, Christina C. Spevak⁴, John Clutterbuck⁵, Vladimir Kapitonov⁶, Jerzy Jurka⁶, Claudio Scazzocchio⁷, Mark Farman⁸, Jonathan Butler¹, Seth Purcell¹, Steve Harris⁹, Gerhard H. Braus¹⁰, Oliver Draht¹⁰, Silke Busch¹⁰, Christophe D'Enfert¹¹, Christiane Bouchier¹², Gustavo H. Goldman¹³, Deborah Bell-Pedersen¹⁴, Sam Griffiths-Jones¹⁵, John H. Doonan¹⁶, Jaehyuk Yu¹⁷, Kay Vienken¹⁸, Arnab Pain¹⁵, Michael Freitag¹⁹, Eric U. Selker¹⁹, David B. Archer²⁰, Miguel Á. Peñalva²¹, Berl R. Oakley²², Michelle Momany²³, Toshihiro Tanaka²⁴, Toshitaka Kumagai²⁵, Kiyoshi Asai²⁵, Masayuki Machida²⁵, William C. Nierman^{2,26}, David W. Denning²⁷, Mark Caddick²⁸, Michael Hynes²⁹, Mathieu Paoletti²⁰, Reinhard Fischer¹⁸, Bruce Miller³⁰, Paul Dyer²⁰, Matthew S. Sachs⁴, Stephen A. Osmani²² & Bruce W. Birren¹

The aspergilli comprise a diverse group of filamentous fungi spanning over 200 million years of evolution. Here we report the genome sequence of the model organism *Aspergillus nidulans*, and a comparative study with *Aspergillus fumigatus*, a serious human pathogen, and *Aspergillus oryzae*, used in the production of sake, miso and soy sauce. Our analysis of genome structure provided a quantitative evaluation of forces driving long-term eukaryotic genome evolution. It also led to an experimentally validated model of mating-type locus evolution, suggesting the potential for sexual reproduction in *A. fumigatus* and *A. oryzae*. Our analysis of sequence conservation revealed over 5,000 non-coding regions actively conserved across all three species. Within these regions, we identified potential functional elements including a previously uncharacterized TPP riboswitch and motifs suggesting regulation in filamentous fungi by Puf family genes. We further obtained comparative and experimental evidence indicating widespread translational regulation by upstream open reading frames. These results enhance our understanding of these widely studied fungi as well as provide new insight into eukaryotic genome evolution and gene regulation.

The aspergilli are a ubiquitous group of filamentous fungi spanning over 200 million years of evolution. Among the over 185 aspergilli are several that have an impact on human health and society, including 20 human pathogens as well as beneficial species used to produce foodstuffs and industrial enzymes¹. Within this genus, *A. nidulans* has a central role as a model organism. In contrast to most aspergilli, *A. nidulans* possesses a well-characterized sexual cycle and thus a well-developed genetics system. Half a century of *A. nidulans*

research has advanced the study of eukaryotic cellular physiology, contributing to our understanding of metabolic regulation, development, cell cycle control, chromatin structure, cytoskeletal function, DNA repair, pH control, morphogenesis, mitochondrial DNA structure and human genetic diseases.

We present here the genome sequence for *A. nidulans*, and a comparative genomics study with two related aspergilli: *A. fumigatus*² and *A. oryzae*³. *A. fumigatus* is a life-threatening human pathogen, and

¹The Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02142, USA. ²The Institute for Genomic Research, Rockville, Maryland 20850, USA. ³Department of Computer Science, Stanford University, Stanford, California 94305, USA. ⁴Department of Environmental and Biomolecular Systems, Oregon Health & Science University, 20 000 NW Walker Road, Beaverton, Oregon 97006-8921, USA. ⁵Division of Molecular Genetics, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G11 6NU, UK. ⁶Genetic Information Research Institute, 1925 Landings Drive, Mountain View, California 94043, USA. ⁷Institut de Génétique et Microbiologie, Institut Universitaire de France, Université Paris-Sud, UMR8621, 91405 Orsay Cedex, France. ⁸Department of Plant Pathology, Plant Science Building, 1405 Veteran's Drive, University of Kentucky, Lexington, Kentucky 40546-0312, USA. ⁹Plant Science Initiative and Department of Plant Pathology, University of Nebraska, Lincoln, Nebraska 68588, USA. ¹⁰Department of Molecular Microbiology and Genetics, Institute for Microbiology and Genetics, Georg-August-University Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany. ¹¹Institut Pasteur, Unité Postulante Biologie et Pathogénicité Fongiques, INRA USC 2019, 75724 Paris Cedex 15, France. ¹²Institut Pasteur, Génopole-PF1, 75724 Paris Cedex 15, France. ¹³Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Brazil. ¹⁴Department of Biology, Texas A&M University, College Station, Texas 77843, USA. ¹⁵The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. ¹⁶John Innes Centre, Colney Lane, Norwich NR4 7UH, UK. ¹⁷Department of Food Microbiology and Toxicology, The University of Wisconsin-Madison, 1925 Willow Drive, Madison, Wisconsin 53706-1187, USA. ¹⁸Max Planck Institute for terrestrial Microbiology, D-35043 Marburg and Institute for Applied Biosciences at the University of Karlsruhe, D-76187 Karlsruhe, Germany. ¹⁹Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA. ²⁰School of Biology, University Park, University of Nottingham, Nottingham NG7 2RD, UK. ²¹Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, Madrid 28040, Spain. ²²Department of Molecular Genetics, The Ohio State University, Columbus, Ohio 43210, USA. ²³The University of Georgia, Department of Plant Biology, 2502 Plant Sciences, Athens, Georgia 30602-7271, USA. ²⁴National Institute of Technology and Evaluation (NITE), 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan. ²⁵National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan. ²⁶The George Washington University School of Medicine, Department of Biochemistry and Molecular Biology, 2300 Eye Street NW, Washington DC 20037, USA. ²⁷Schools of Medicine and Biological Sciences, The University of Manchester, Stopford Building, Manchester M23 9PL, UK. ²⁸Plant Science and Fungal Molecular Biology Research Group, School of Biological Sciences, Donnan Labs, The University of Liverpool, Liverpool L69 7ZD, UK. ²⁹Department of Genetics, University of Melbourne, Parkville, Victoria 3010, Australia. ³⁰Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, Moscow, Idaho 83844-3052, USA.

A. oryzae is used in the production of sake, miso and soy sauce. *A. oryzae* and *A. fumigatus* lack known sexual cycles, and their study relies on *A. nidulans* as a genetic model. Our analysis of these organisms focused on the genomic bases of their differing physiologies, while investigating their common eukaryotic biology. Our results yield new insights into eukaryotic genome evolution, the evolution of mating-type loci, the potential for sexual reproduction in the two asexual species, and the role of conserved sequence elements in gene regulation.

Genome assembly and annotation

The genome sequence of *A. nidulans* was assembled from deep whole-genome shotgun (WGS) coverage obtained by paired-end sequencing from a variety of clone types (see Methods). An average of 13× sequence coverage was generated including ×3 coverage produced and provided by Monsanto. The Arachne package (<http://www.broad.mit.edu/wga/>) was used to assemble the sequence, and the resulting assembly consists of 248 sequence contigs with an N50 length of 282 kilobases (kb) (that is, 50% of all bases are contained in contigs of at least 282 kb). Contigs were assembled into 89 scaffolds with a total length of 30.06 megabases (Mb) (including gaps between contigs) and an N50 length of 2.44 Mb. A total of 28.5 Mb (95%) of the assembly was anchored to the *A. nidulans* genetic map^{4,5} through meiotically mapped markers with sequence and markers located by haploidization or hybridization to electrophoretically separated chromosomes (see Supplementary Information). By comparison with previously published pulse-field gel electrophoresis data, we estimate that the assembly comprises 96.3% of the complete genome.

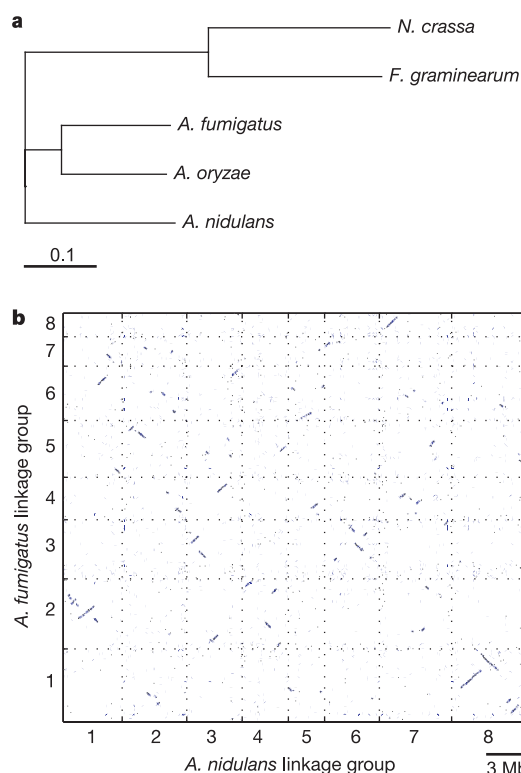


Figure 1 | Phylogenetic tree and representative dot plot. **a**, Phylogenetic tree showing the relationship between three *Aspergillus* species compared using *N. crassa* and *F. graminearum* as an outgroup. Branch lengths correspond to substitutions per site calculated using a maximum likelihood approach. An identical topology was predicted using *C. immitis* as an outgroup. **b**, Dot plot of *A. nidulans* (horizontal) and *A. fumigatus* (vertical) genomes. Axes represent the concatenation of all chromosomes for the corresponding genome. Gridlines indicate the boundaries between chromosomes and axis labels indicate chromosome number. Elements in the dot plot represent protein homology translated to genomic coordinates.

The assembly was annotated using the Calhoun system, as described in the Methods and Supplementary Information.

Phylogenetic relationship

Previous work based on large subunit rDNA data has led to a widely accepted phylogeny of the aspergilli in which *A. nidulans* and *A. oryzae* are more related to one another than *A. fumigatus*⁶. However, single gene phylogenies can contradict organismal phylogenies⁷. In principle, whole-genome data provide greater resolving power by allowing trees to be constructed based on concatenated sets of genes⁷. Using this approach to study the relationship of the three aspergilli, we find support for an alternative phylogeny⁸ (Fig. 1a).

We established this relationship using a set of 3,034 predicted orthologues across the three aspergilli, *Neurospora crassa* and *Fusarium graminearum*. We constructed trees for 75 randomly selected sets of 20 concatenated genes⁷, using the *N. crassa* and *F. graminearum* genes to root the trees (see Methods). All 75 cases produced the phylogeny shown in Fig. 1a in which *A. fumigatus* and *A. oryzae* are sister taxa and *A. nidulans* branches earlier. This phylogeny is further supported by 86% of trees built for each of the 3,034 orthologues individually. Consistent with this phylogeny, *A. fumigatus* has over twice as many genes with top Blast hits to *A. oryzae* than to *A. nidulans*, and *A. oryzae* has almost twice as many genes closer to *A. fumigatus* than *A. nidulans*. *A. nidulans* has roughly a similar number of top hits to *A. fumigatus* and *A. oryzae*. To confirm further the rooting of the tree, we repeated the analysis using predicted gene fragments (see Methods) from the genome sequence of *Coccidioides immitis* as an outgroup (which is closer to the aspergilli than *N. crassa* and *F. graminearum*). Ninety-four per cent (34 out of 36) of 50-gene phylogenies with *C. immitis* as the outgroup support the relationship of Fig. 1a, as do 60.8% (93 out of 153) of single gene phylogenies (only 21% support the rDNA phylogeny).

Overall genome and proteome comparison

Although in the same genus, the three aspergilli differ considerably in their genome sequences. Predicted orthologues shared by all three species (three-way orthologues) display an average of only 68% amino acid identity. *A. fumigatus* and *A. oryzae* share 70% identity, and each has 66–67% identity with *A. nidulans*. This protein identity is comparable to that between mammals and fish⁹, which diverged ~450 million years ago. The three species also differ considerably in genome size (Table 1). The largest, *A. oryzae* (36 Mb), is 31% bigger than the smallest, *A. fumigatus* (28 Mb), and 24% bigger than *A. nidulans* (30 Mb). This difference seems to be due to an acquisition of sequence in *A. oryzae*³ rather than loss in both *A. nidulans* and *A. fumigatus*. Finally, the genomes show extensive structural reorganization (Fig. 1b).

Conserved synteny and genome evolution

These three aspergilli provide an opportunity to study eukaryotic genome evolution over a divergence approaching the limit of conserved long-range synteny. To characterize pairwise conserved synteny, we used an algorithm based on hierarchical clustering that delineates regions of conserved synteny while also retaining information about internal micro-rearrangements (see Methods). Using this method, the majority (77–79%) of each genome assembly could be mapped to conserved syntenic blocks with at least one other genome (Table 2). Figure 2 shows a projection of the homologous blocks onto the chromosomes of *A. nidulans* and, contrasted with Fig. 1b, illustrates the considerable extent of conserved synteny despite extensive rearrangement.

The results of this analysis reveal two notable trends. First, large regions lacking detectable long-range synteny are readily apparent. As has been observed for mammals, nematodes and yeasts¹⁰, repeats and subtelomeric sequences are associated with these heavily rearranged regions. This may have specific implications for fungi, as subtelomeric regions in the aspergilli are enriched for secondary

Table 1 | Comparison of genome characteristics

Genome characteristic	<i>A. nidulans</i>	<i>A. fumigatus</i>	<i>A. oryzae</i>
General			
Assembly size (bp)	30,068,514	27,980,910	37,047,050
G+C (%)	50	49	48
Protein coding genes	9,541	9,926	14,063
Protein coding genes >100 amino acids	9,396	9,009	12,074
Predicted protein coding sequences >100 amino acids			
Coding (%)	50	49	45
Gene density (1 gene every <i>n</i> bp)	3,151	2,938	2,613
Median gene length (mean)	1,547 (1,868)	1,389 (1,644)	1,152 (1,414)
Average number of exons per gene	3.6	2.8	2.9

metabolite genes thought to have a role in niche adaptation and virulence. The rapid rearrangement of subtelomeric regions may facilitate the species-specific evolution of these genes (Supplementary Information).

The second notable trend is the distribution of lengths of unbroken regions between micro-rearrangements within pairwise syntenic blocks. The random breakage model of genome evolution predicts that such lengths should be exponentially distributed. Although the mean breakpoint lengths differ, in all three pairwise comparisons the distribution of lengths shows close agreement with the model prediction (Supplementary Information). It thus seems that syntenic blocks, comprising the majority of the *Aspergillus* chromosomes, are evolving in a manner consistent with random breakage.

For each pairwise comparison, the third *Aspergillus* genome allows the determination of rearrangements specific to each branch of the unrooted tree (see Methods). The results of this analysis provide a quantitative estimate of the different rearrangements contributing to long-term eukaryotic genome evolution (Fig. 3).

Structural evolution not correlated with molecular evolution

In vertebrates, nematodes and arthropods, it has been reported that the rates of structural evolution and nucleotide evolution are correlated^{11–13}. However, our analysis of the *Aspergillus* genomes suggests that this expected correlation does not always hold for eukaryotes.

The data in Fig. 3 reveal a considerably higher overall rate of genome reorganization in the lineage of *A. oryzae* compared to *A. fumigatus*. Nearly all categories of disruption are at least twofold greater in *A. oryzae* relative to *A. fumigatus*. For example, *A. oryzae* displays a more than twofold greater rate of insertion than *A. fumigatus*. This is consistent with the larger genome size of *A. oryzae*³. Surprisingly, our analysis also indicates that chromosomal breaks are more common in *A. oryzae* than *A. fumigatus*. Although apparent intrachromosomal rearrangements could arise from successive inversion events, this cannot explain interchromosomal rearrangements. These interchromosomal breaks are also not the result of assembly error, as confirmed by optical mapping³ and polymerase chain reaction (PCR) validation of eight predicted interchromosomal breaks.

In contrast, several measures indicate that the rates of amino acid evolution in predicted orthologues are similar between these two

species. An examination of predicted three-way orthologues shows that the distribution of amino acid identity is roughly similar for both *A. oryzae* and *A. fumigatus* compared to *A. nidulans*, as are non-synonymous divergences (Supplementary Information). In addition, branch lengths predicted from phylogenetic trees (see above) indicate a comparable rate of substitution for both *A. oryzae* and *A. fumigatus*. Taken together, these data lead to the conclusion that structural and molecular evolution in the aspergilli is not correlated. A similar conclusion has been reached in the analysis of two microsporidian genomes, although in this case gene evolution seems to be accelerated relative to genome rearrangement¹⁴. Thus, large-scale and small-scale evolutionary processes in eukaryotes can operate at different relative rates in a species-specific manner.

Sex and the evolution of the mating-type loci

Unlike *A. nidulans*, which has a known sexual cycle, *A. fumigatus* and *A. oryzae* are only known to reproduce through asexual mitotic spores. We sought insight into the evolution of this apparent difference by comparing the three genomes. Our results, in conjunction with an accompanying paper and another study^{2,15}, suggest that both *A. fumigatus* and *A. oryzae* may be capable of sexual reproduction.

Sexual reproduction in ascomycete filamentous fungi is governed, in part, by two different mating-type genes that establish sexual compatibility: one gene encodes a protein with a high mobility group (HMG) domain, whereas the other encodes a protein with an alpha box domain. We refer to these genes here as the HMG and alpha mating-type genes, and to their chromosomal locations as MAT loci. Homothallic fungi typically possess both mating-type genes and are self-fertile. Heterothallic fungi possess only one mating-type gene and require a partner with a different mating-type gene. In heterothallics, the two mating-type genes typically occupy the same chromosomal location in different haploid genomes but lack sequence similarity, and are thus termed idiomorphs rather than alleles¹⁶.

A. nidulans is known to be homothallic, and both HMG and alpha mating-type genes have been identified^{17,18}. Our analysis confirmed that the HMG and alpha loci are unlinked, which is unusual although not unprecedented in homothallic fungi¹⁹. We identified a single HMG mating-type gene in *A. fumigatus*, as previously reported²⁰, and a single alpha mating-type gene in *A. oryzae*.

A comparison of all four MAT loci revealed extensive conserved

Table 2 | Characteristics of pairwise conserved synteny

Reference	Coverage (Mb) (percentage of reference)*				Maximum/mean block length (kb)‡		
	<i>A. nidulans</i>	<i>A. fumigatus</i>	<i>A. oryzae</i>	Either†	<i>A. nidulans</i>	<i>A. fumigatus</i>	<i>A. oryzae</i>
<i>A. nidulans</i>	–	20.5 (68)	20.4 (68)	21.6 (72)	–	175	114
<i>A. fumigatus</i>	20.4 (73)	–	20.7 (74)	21.5 (77)	2,429	–	168
<i>A. oryzae</i>	23.3 (63)	24.3 (66)	–	25.4 (69)	943	1,159	–

* Coverage of reference organism assembly by pairwise conserved syntenic blocks (>10 kb in length) to each target genome.

† Coverage of reference organism assembly by pairwise conserved syntenic blocks (>10 kb in length) in either other genome.

‡ Upper right half shows mean block sizes and lower left half shows maximum sizes across all blocks using either organism as reference.

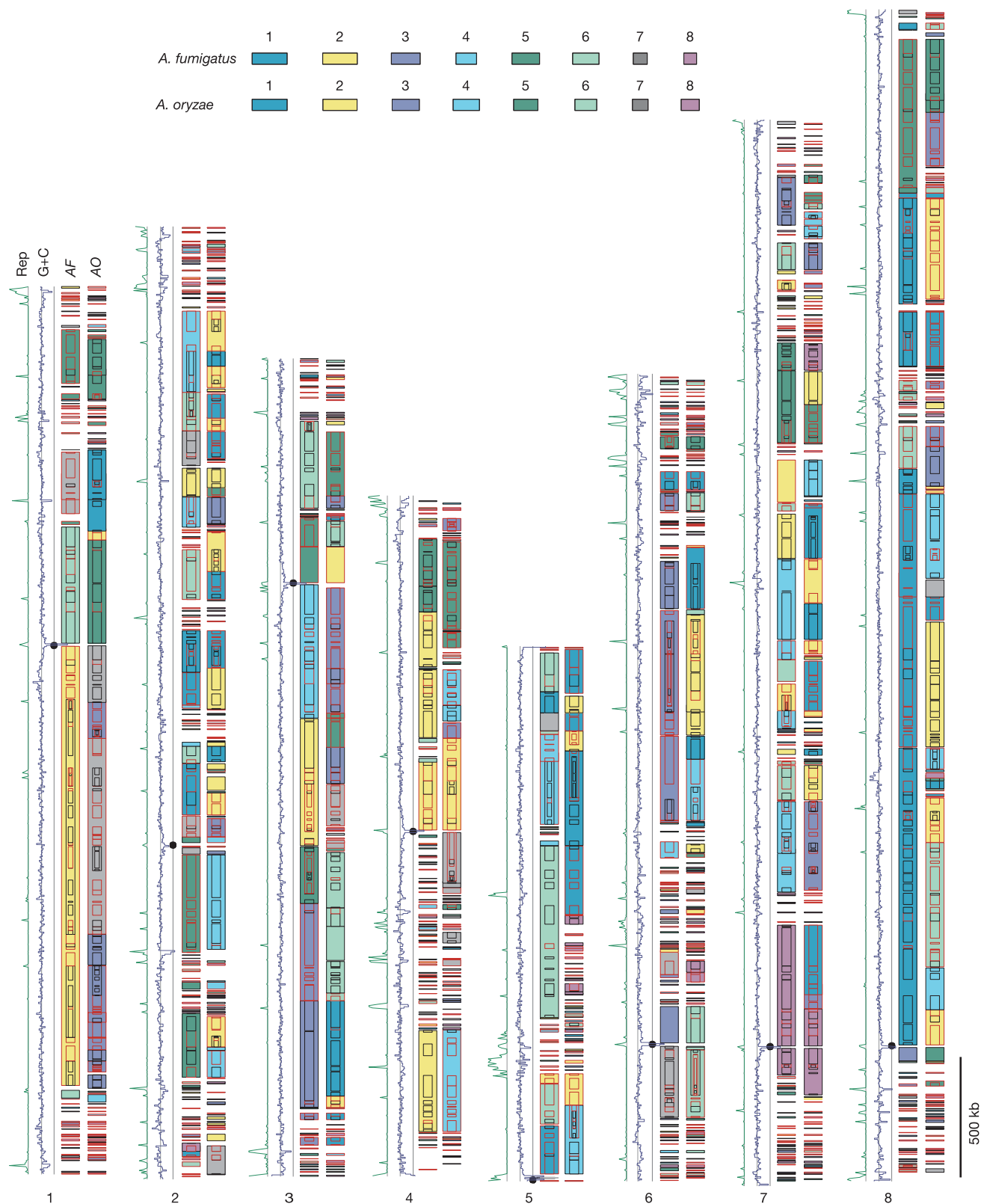


Figure 2 | *Aspergillus* comparative map. Conserved synteny between *A. nidulans* and *A. oryzae* and *A. fumigatus*. Syntenic regions are represented by two vertical columns of coloured blocks. The left and right columns represent syntenic blocks to *A. fumigatus* and *A. oryzae*, respectively, coloured by chromosome as indicated by the key. Nested blocks show synteny at finer resolutions. Blocks outlined in black are in opposite

orientations in *A. nidulans* relative to those in red. Red blocks in black blocks (and vice versa) represent inversions. The green and purple lines display repeat density (Rep) and G+C content (G+C) in *A. nidulans*, both in 5-kb windows with increasing values to the left. Black circles represent centromeres.

synteny (Fig. 4a). The *A. oryzae* alpha locus and the *A. fumigatus* HMG locus display conserved synteny over 1.7 Mb on either side of the mating-type genes. Within this region of conserved synteny, the two mating-type genes occupy nearly identical positions, although offset with different orientations. Furthermore, one flank of both the *A. fumigatus* and *A. oryzae* loci is syntenic with 409 kb of the *A. nidulans* HMG locus downstream region, whereas the other flank is syntenic with 34 kb of the *A. nidulans* alpha locus downstream region (Fig. 4a). The four loci also show conservation of a number of genes associated with MAT loci in other species including *N. crassa* or one of nine yeast species previously analysed²¹.

Extending the analysis to 215 genes implicated in the fungal mating process, pheromone response, meiosis and fruiting body development revealed that every gene (except for the mating-type genes) that can be identified in *A. nidulans* is also present in both *A. fumigatus* and *A. oryzae* (Supplementary Information), including several genes for which the only known function is related to sexual reproduction.

Although sexual reproduction may have been lost very recently in both *A. fumigatus* and *A. oryzae*, providing one explanation for the residual presence of mating process genes, these data suggested the possibility that both *A. fumigatus* and *A. oryzae* may be capable of sexual reproduction. Moreover, the pattern of synteny among the four MAT loci leads to an evolutionary scenario for this hypothesis, as shown in Fig. 4b. According to this model, it is predicted that *A. oryzae* and *A. fumigatus* isolates exist with the opposite mating-type genes to those present in the strains that were sequenced. In addition, these opposite mating-type genes should be present at the identical locus, consistent with a heterothallic idiomorphic configuration.

As reported in detail in another study¹⁵, these predictions have been experimentally verified. Using a PCR-based multiplex mating-type assay, isolates of both mating types of *A. fumigatus* and *A. oryzae* were identified. For both species, the opposite MAT locus from the complete genome was sequenced and demonstrated to have the idiomorphic organization predicted. Within the idiomorphic region the opposite mating-type genes appear to be offset with respect to one another, as predicted by our model. In addition, the *A. fumigatus* alpha MAT locus was found to contain a 360-base pair (bp) fragment of an HMG gene¹⁵ neighbouring the idiomorphic region, suggesting that the transition from homothallism to heterothallism in the *A. oryzae* and *A. fumigatus* ancestor occurred by gene loss (Fig. 4b).

Although the model of Fig. 4b predicts a homothallic ancestor for all three species, it is possible that heterothallism was ancestral and a transition to homothallism occurred in the *A. nidulans* lineage. This would be consistent with data from *Cochliobolus* species for which heterothallism appears to be ancestral, and conversions to homothallism have been described¹⁹. However, two factors conflict with this scenario for the aspergilli. First, the offset positions of the mating-type genes within the idiomorphic regions of the *A. fumigatus* and *A. oryzae* MAT loci, and the apparent fragment of the HMG gene neighbouring the *A. fumigatus* alpha locus, are consistent with gene loss from a homothallic ancestor. Second, heterothallism in the aspergilli is rare^{22,23}. Only three heterothallic aspergilli have been previously characterized, of which one, *A. heterothallicus*, groups in phylogenies with known homothallic species, suggesting a conversion to heterothallism in this case as well²². Mitotic, homothallic and heterothallic species are observed intermixed in several fungal lineages, leading to debates about the fungal ancestral state^{19,24–26}. Taken together, our results provide evidence that conversion from homothallism to heterothallism is possible, and suggest that the predominance of a particular sexual strategy may vary within different clades.

Although the finding of MAT genes in supposedly asexual fungi has been previously reported^{15,27–32} and genes related to sexual reproduction have been found in the 'asexual' yeast *Candida albicans*, this report is the first comprehensive survey of sexual reproduction

genes in two different filamentous fungi thought to be asexual. In addition, our results provide an experimentally supported evolutionary model associating large-scale synteny and genome rearrangement with a specific and significant difference in biology between these aspergilli. These results for *A. fumigatus* and *A. oryzae* have important and specific potential implications for health and industry. The lack of a sexual cycle in *A. fumigatus* and *A. oryzae* has precluded classical genetic analysis, impeding efforts to study these organisms and necessitating the use of the relatively distant *A. nidulans* as a genetic model. The possibility for mating—still speculative at this stage—raises the medically and industrially important potential for developing genetic tools for these fungi.

Conserved non-coding sequences

Detecting and characterizing conservation of sequences outside of protein coding regions is a promising method for identifying potential functional elements. Regulation in yeast has been extensively studied; however, in the aspergilli few transcription factor binding sites have been experimentally verified. Comparing the three aspergilli provides an opportunity to identify the most constrained functional elements.

To do so we aligned three-way orthologous genes including 1 kb of sequence upstream and downstream using Mlagan³³. Strict filters were then applied to delineate unambiguous orthologous intergenic regions (see Methods). Given the divergence of the aspergilli, it is expected that intergenic regions would not show significant conservation, and frequently this was found to be the case. However, in many instances, blocks of nearly perfect three-way conservation

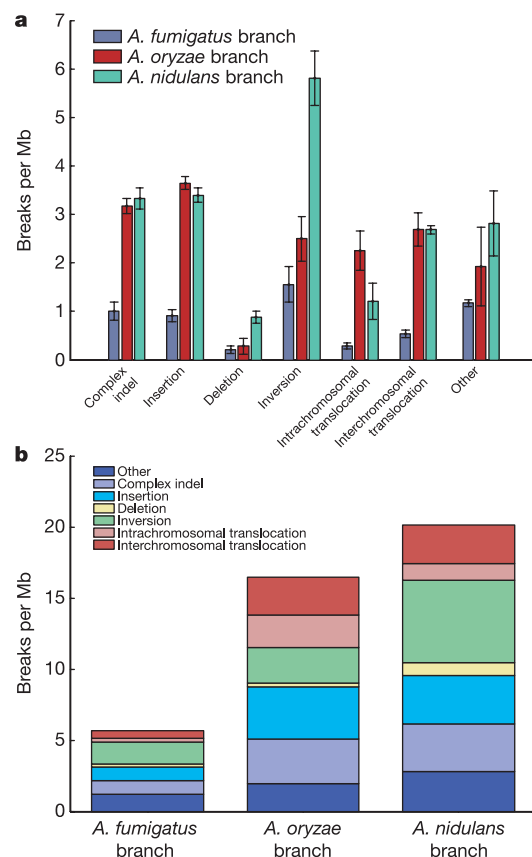


Figure 3 | Rates of branch-specific rearrangements. **a**, The rates of different breaks broken down by break type for each branch. Bars represent minimum and maximum values obtained using either of the two non-target genomes as reference (see Methods). **b**, A stacked plot of the same data showing the relative contribution of break types within each branch for all three branches. See text and Methods for more details.

are observed. An example for one intergenic region is shown in Fig. 5a.

To assess which regions were conserved owing to purifying selection rather than neutral mutation or chance, we developed models

for alignments of neutral and random sequence. Unlike mammals, where ancient conserved repeats provide a natural model for neutral evolution, few such repeats exist in the aspergilli (see below). Instead, we synthesized alignments of neutral sequence by concatenating

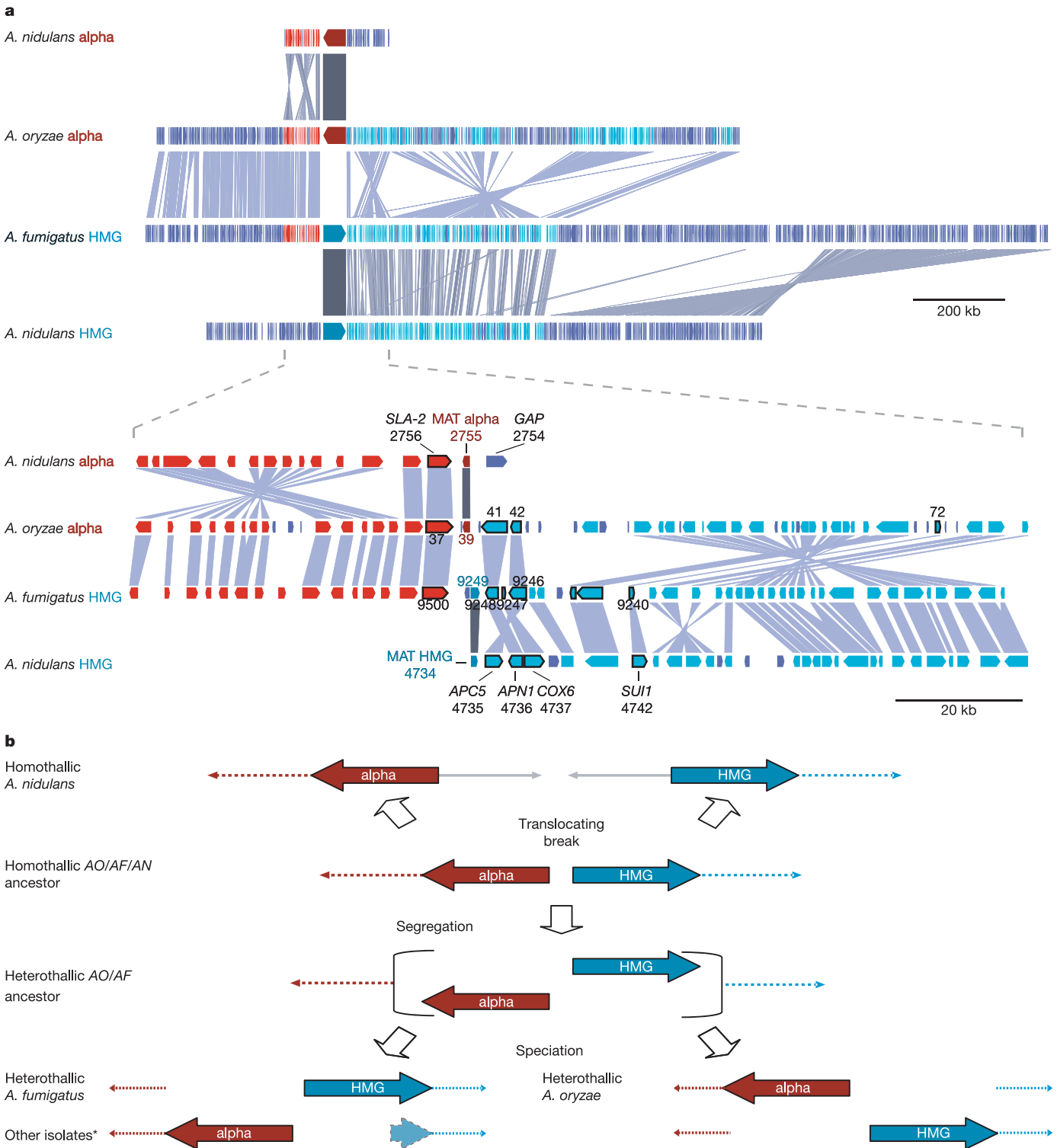


Figure 4 | Comparison and evolutionary model of *Aspergillus* MAT loci. **a**, Conserved synteny between loci. Grey lines indicate predicted orthologues. Red genes indicate orthologues from the left flank (as drawn) of the *A. nidulans* alpha locus with the left flanks of the *A. fumigatus* and *A. oryzae* loci. Cyan genes indicate orthologues with the right flank of the *A. nidulans* HMG locus. The bottom panel shows the region near mating-type genes. Genes labelled and outlined in black are associated with MAT loci in other fungi. Only partial accession numbers (suffixes) are shown in the

figure. For full accession numbers, the numbers shown in the panel should replace the asterisks in the following examples: *A. nidulans* (AN****.1); *A. fumigatus* (59.m0****); *A. oryzae* (AO0703270000**). **b**, Model of structural evolution of the MAT loci. Braces represent multiple haplotypes at the same genomic locus. The experimental identification of other isolates (indicated by an asterisk) was reported in another study¹⁵. The light blue arrow indicates a 360-bp HMG gene fragment. AF, *A. fumigatus*; AN, *A. nidulans*; AO, *A. oryzae*.

randomly selected aligned columns of fourfold degenerate sites. We also controlled for chance alignment introduced by potential aligner bias by aligning randomly selected intergenic regions. Using a simple conservation scoring function we calculated maximal scoring subsequences and compared results between orthologous regions and the control models (Fig. 5b). A noteworthy aspect of the data in Fig. 5b is the similarity between the neutral and random models. According to our models, neutral sequence is effectively saturated for mutation, confirmed by an independent analysis of synonymous sites in protein coding sequences.

Comparing results between real and control alignments, we selected a minimum score of 22 for regions unlikely ($P < 0.015$) to be conserved by neutral evolution or chance. We denote a subsequence scoring above this threshold as a high-scoring conserved sequence (HCS). On the basis of this analysis, we predict 5,801 HCSs corresponding to ~2% of alignable orthologous intergenic regions.

Prediction of functional motifs

We expect HCSs to be enriched for functional elements. The challenge is to discover these functional elements and make testable predictions about their biological functions. In preliminary analyses, several conserved regions could be identified as known functional elements. For example, we observed conservation delimiting a known 3' untranslated region (UTR) element of the *A. nidulans* *areA* gene that regulates messenger RNA stability in response to cellular nitrogen levels³⁴. We also identified three TPP binding riboswitches, one of which has not been described in *Aspergillus* (Supplementary Information).

To enrich computationally HCSs for sequences corresponding to functional elements, and to derive clues about their biological functions, we modified the approach used by ref. 35 (see Methods). Briefly, we identified common subsequences (or 'patterns') that appeared in at least four HCSs across all three *Aspergillus* genomes. These patterns were searched in three-way conserved orthologues to identify genes in which the subsequence occurred in the 500-bp upstream or downstream regions (a 'co-occurrence'). A number of conservation criteria were then applied (see Methods). We identified a total of 69 conserved patterns ('conpats'), occurring in at least four HCSs, that showed enrichment for co-occurrences and exhibited a bias for occurring 500 bp upstream or downstream of genes.

The results of this analysis for the 35 most common patterns are shown in Fig. 6 (all 69 patterns available in Supplementary Information). These include motifs that match known or predicted *Aspergillus* or other fungal functional sequences. For example, CPC/GCN4, the master regulator of the cross pathway control system in fungi, is known to bind to the palindromic site TGASTCA³⁶. In yeast, microarray studies have identified 539 genes probably regulated by GCN4 that show a preference for amino acid biosynthetic genes and several ribosomal proteins and translation factors³⁷. One of the patterns identified by our analysis (ID 2483) matched the CPC/GCN4 binding site, co-occurred preferentially upstream of genes, and was enriched in genes associated with amino acid transport and metabolism (COG category E), and translation, ribosomal structure and biogenesis (category J). Furthermore, the 19 genes with co-occurrences of this pattern include 7 (37%) predicted orthologues to the 539 known yeast regulated genes,

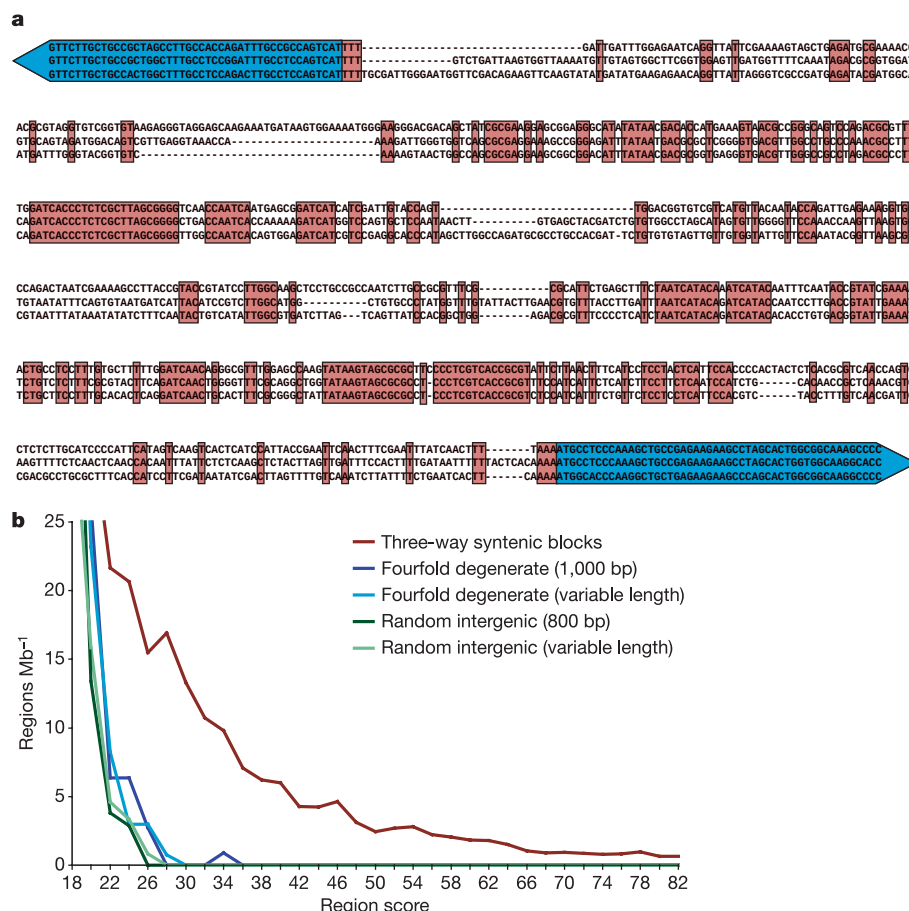


Figure 5 | Active conservation of non-coding regions. **a**, Example region between a conserved pair of orthologous histone H2A and H2B genes (left and right blue arrows). The three lines from top to bottom correspond to the sequences of *A. nidulans*, *A. fumigatus* and *A. oryzae* aligned using Mlagan.

Letters on the red background indicate 100% conserved bases.

b, Conservation scores of maximal subsequences for observed intergenic alignments (red), and models of neutral and random sequence alignment (both fixed and variable length).

representing a sixfold enrichment (P -value $< 1 \times 10^{-5}$). This includes two known *A. nidulans* CPCA regulated genes (*trpB*³⁸ and *hisH*³⁹). This pattern probably corresponds to the known *Aspergillus* CPCA binding site.

A second pattern shows strong correspondence with the binding site for Puf family genes⁴⁰. Puf proteins regulate mRNA translation and mRNA decay through interactions with 3' UTR sequences. In *Saccharomyces cerevisiae*, which has five Puf genes, Puf3p has been shown to bind specifically mRNAs encoding mitochondrial proteins⁴⁰ and requires a 3' UTR motif with consensus UGUANAUA⁴⁰. Four different patterns identified by our analysis (ID 1710, 2077, 1144 and 2378; see the full table in Supplementary Information) match or include the Puf binding motif and display a strong downstream bias. Three also show enrichment for predicted orthologues in *S. cerevisiae* that localize to mitochondria. Taking all four patterns together, we find a 6.8-fold enrichment ($P < 4.3 \times 10^{-11}$) for genes with orthologues to yeast mitochondrial genes. In addition, we find a threefold enrichment ($P < 0.0003$) for genes with yeast orthologues predicted to be bound by Puf genes in a genome-wide affinity tag analysis⁴⁰. Although a functional role for the Puf family has not

been experimentally demonstrated in *Aspergillus*, all three genomes possess five loci with 5–8 Puf domains, as predicted by HMMER and PFAM (including one with a predicted RNA binding domain, as with Puf1p and Puf2p in yeast). Together these data suggest that, as in yeast, Puf genes may bind to and regulate mitochondrial mRNAs in the aspergilli.

Only a small number of transcription factor binding sites and control elements are known for filamentous fungi in general, including *Aspergillus*. These predicted patterns are thus promising targets for future experimental validation.

Regulatory upstream open reading frames

A significant proportion (32%) of HCSs are conservatively predicted to lie within transcribed but untranslated regions of genes (UTRs), consistent with the known role of UTRs in regulating gene expression, particularly mRNA translation (for example, Puf binding domains). One important class of translational control elements is short upstream open reading frames (uORFs) in 5' UTRs⁴¹, which can regulate the expression of downstream protein-coding genes in several ways. First, they can modulate the efficiency of ribosome

ID	Conpat consensus	Similar sequence	No. 5' sites	No. 3' sites	5' / 3' bias?	Aligned (%)	Strand bias?	Enriched in COG categories (no. genes in category)
1821	TCACGTG	bHLH	115	8	up***	35	Yes***	J: Translation, ribosomal structure, biogenesis (16)*
1622	ATCTTATC	<i>areA/gln3</i>	25	3	up*	60		
234	gTTTg... TgTTTg		46	11	up**	35		A: RNA processing, modification (8)
1934	TCACATGA	bHLH	36	3	up***	47	Yes*	I: Lipid transport, metabolism (10)**
1710	TgTATATg	Puf	5	20	down*	30	Yes***	
896	TGGgCcGTCG		14	2	up	43	Yes	
1602	ACCGCCT		34	0	up***	50		
1879	cTTATCGAT		24	1	up**	63	Yes	
1132	CGCCTCT		14	2	up	36	Yes***	
2683	TCTCCGC		27	1	up***	48	Yes*	
2223	GGCGTT		5	16	down	63	Yes***	
1291	TgAGTcAG	yAP1	21	1	up**	62	Yes	
992	GCATAGCg		1	21	down***	86	Yes***	
2611	TGTACAT		0	17	down***	59		
1739	GGGgCTAGGG		9	0	up	44	Yes*	J: Translation, ribosomal structure, biogenesis (4)
418	cAAAc...TcCAA		33	1	up***	45		
2077	TgTAcAAATA	Puf	1	9	down	56	Yes	
973	cACCGCCT		74	6	up***	36	Yes*	A: RNA processing, modification (12)*
2488	cACCGCG		37	4	up***	38	Yes**	T: Signal transduction mechanisms (11)**
1144	TGTACTAT	Puf	2	11	down	36	Yes***	
1635	cGACAAACc		13	1	up	69	Yes***	J: Translation, ribosomal structure, biogenesis (8)***
1940	cTgGCGTT		2	19	down**	47	Yes***	D: Cell cycle/division, chromosomal partitioning (5)*
2199	CCACGTGC	bHLH	40	3	up***	30		
1426	AcTCGG		14	1	up	50		C: Energy production, conversion (5)*
2525	GACGCGT	<i>stuA/mbp1</i>	18	1	up*	39	Yes	L: Replication, recombination, repair (10)***
813	ACGTCAg	yATF-B	11	1	up	45		
1650	GAAA-TT		11	0	up	73		J: Translation, ribosomal structure, biogenesis (4)
2093	CCCCcCA		44	5	up***	30	Yes***	I: Lipid transport, metabolism (8)
1998	CCTCGG-A		25	1	up***	40	Yes	I: Lipid transport, metabolism (7)*
2497	TGCA-AG		0	8	down	100	Yes**	
751	AcCGCCT-c		25	2	up**	40		
2483	TGACTCA	<i>cpcA/gcn4</i>	19	4	up	26		E: Amino acid transport, metabolism (7)*
			36	0	up***	25		J: Translation, ribosomal structure, biogenesis (7)**
2249	TTTTTTT							A: RNA processing, modification (7)
								J: Translation, ribosomal structure, biogenesis (13)***
1059	TCGG-CCG		12	1	up	92		
2451	GATATCc		0	10	down*	80		

Figure 6 | Selected conserved patterns. Column one shows the conpat unique ID. Column two shows the sequence logo representation of conpat weight matrix. Column three shows fungal binding factors with sequence similarity to the conpat. Columns four and five show the number of genes with a co-occurrence of conpat upstream and downstream. Column six shows the preference for co-occurring preferentially 5' or 3' of the gene. Column seven shows the fraction of co-occurrences overlapping three-way

conserved regions. Column eight shows the preference for co-occurring on a particular strand relative to the gene. Column nine shows COG categories showing significant enrichment (the number of genes with co-occurring conpats in the category is indicated in parentheses). Enrichment results for yeast orthologue cellular location are available in the Supplementary Information. Single asterisk, $P < 1 \times 10^{-3}$; double asterisk, $P < 1 \times 10^{-4}$; triple asterisk, $P < 1 \times 10^{-5}$. bHLH, basic helix-loop-helix.

re-initiation at downstream start codons in a manner dependent on cellular state. Second, uORFs can produce *cis*-acting peptides that stall ribosomes. Finally, the presence of uORFs can affect mRNA stability. Functional uORFs can be as short as three amino acids and occur at varying distances and multiplicity upstream of the protein-coding gene, occasionally overlapping the downstream start codon. Functional uORFs have been reported in a range of species including plants, animals and fungi⁴¹. In *Aspergillus*, a small number of genes with validated uORFs have been reported (Supplementary Information). To determine the full extent with which they may regulate gene expression, we analysed the three *Aspergillus* genomes for uORFs.

We identified UTR sequences using expressed sequence tag (EST) alignments for *A. nidulans* genes, and examined them for open reading frames (see Methods). Of 1,606 genes with identified UTRs, 21% (358) have upstream ORFs. A similar proportion was found (18% or 82 out of 463) when we restricted our analysis to three-way orthologues for which the start codons align exactly within multiple alignments, suggesting that these uORFs are not due to mis-annotation. A corresponding analysis of genes with ESTs in the *N. crassa*, *F. graminearum* and *Magnaporthe grisea* genomes found uORFs associated with 22%, 10% and 16% of genes, respectively. We further extended the analysis in *A. nidulans* using a conservative estimate of 5' UTR length (see Methods), and identified an additional 958 genes with potential uORFs of which 165 genes have three-way orthologues. In total, 1,316 genes in *A. nidulans* are predicted to possess uORFs.

Not all identified uORFs have a detectable impact on gene expression⁴¹. To enrich the set of predicted uORFs for those likely to be functional, we looked for those conserved in all three aspergilli. On the basis of a strict criterion requiring alignment of the uORF start and stop codons, we find 38 conserved uORFs (13% of 331 genes with uORFs and predicted orthologues) (Supplementary Information). Of these corresponding *Aspergillus* genes, 14 have predicted uORFs upstream of orthologues in *N. crassa*, *F. graminearum* or *M. grisea*. Additionally, three also have predicted orthologues in *S. cerevisiae* with uORFs conserved across four related yeast species.

These 38 conserved uORFs represent strong candidates for experimental investigation. As a preliminary validation, we tested two novel conserved uORFs for their ability to modulate protein synthesis *in vitro* (Fig. 7). Briefly, oligonucleotides containing each uORF were fused to a luciferase reporter gene, and controls were constructed with disabled uORF and/or reporter gene start codons. Differential expression in a cell-free translation system between intact and control constructs measures the impact of the uORF on translation. This system can detect small (twofold) changes in translation, and can discriminate uORFs that do not reduce translation *in vivo* from those that do (see Methods). As can be seen in Fig. 7, both uORFs tested display a 5–10-fold repressive effect on the translation of the downstream reporter gene.

These results provide the first genome-wide list of predicted conserved uORFs for any organism, and suggest that uORFs could have a substantial role in regulating gene expression in *Aspergillus*. Previous reports estimate that 2–4% of genes in *S. cerevisiae* contain uORFs⁴², whereas a review of sequences in UTRdb predicted that 5–10% of eukaryotic UTRs contain ORFs⁴³. Our results predict that in filamentous fungi the proportion may be twice as high.

Aspergillus physiology

Peroxisomes are organelles containing enzymes for the breakdown of fatty acids (β -oxidation), removal of hydrogen peroxide and synthesis of cholesterol and bile acids. Peroxisomes have critical roles in fungi where they are involved in growth, secondary metabolism and pathogenesis. In mammals, defects can lead to developmental and neurological disorders. Proteins are targeted to the peroxisome either by a carboxy-terminal tripeptide sequence or an amino-terminal nine-amino-acid sequence. Using these signals we predicted

peroxisomal proteins in *Aspergillus* (Supplementary Information). Our analysis reveals peroxisomes in *Aspergillus* to be more similar to mammals than yeasts in two respects. First, our data suggest that the aspergilli, like mammalian cells, perform β -oxidation in both peroxisomes and mitochondria and possess two sets of genes for all β -oxidation enzymes targeted to both the mitochondria and the peroxisome, as supported by recent experimental results⁴⁴. In contrast, *S. cerevisiae* metabolizes fatty acids fully to acetyl-CoA only in peroxisomes⁴⁵. Second, *Aspergillus* peroxisomes are more similar to those of mammals than those of yeasts in that they possess putative peroxisomal acyl-CoA dehydrogenases. In addition, all three aspergilli appear to encode both mitochondrial and peroxisome forms of an ATP-dependent protease of the LON (La domain) family associated with peroxisomes in mammals⁴⁶. *S. cerevisiae* has a single copy of this protease (*Pim1/Lon1*) targeted to mitochondria⁴⁷.

One of the hallmarks of the filamentous fungi is their ability to undergo polarized hyphal growth. This requires positional cues that mark polarized growth sites, locally activating Rho-related GTPase signalling modules that promote cytoskeletal reorganization⁴⁸. The three aspergilli possess the expected genes involved in signalling and cytoskeletal organization for polarized growth, but there is a marked lack of known positional markers (such as the yeast bud site markers Bud3p, Bud8p and Bud9p; see Supplementary Information). Proteins implicated in the transport or modification of bud markers, including Axl1p, Rax1p and Bud7p, were predicted, however. This suggests that filamentous fungi mark polarized growth sites with positional cues, but that the markers themselves may consist of novel cell wall proteins.

Most of the interspersed repeats in all three genome sequences correspond to relics of transposable elements (see Supplementary Information). Surprisingly, only 1.3% of the largest genome

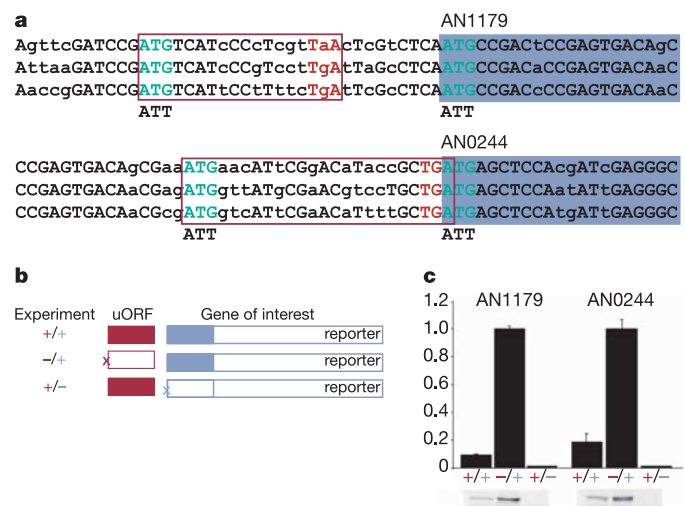


Figure 7 | Prediction and validation of conserved uORFs. Conserved uORFs show a 5–10 times repressive effect on reporter gene translation.

a, Alignments of two tested uORFs. uORFs are shown in purple boxes, and protein-coding genes in a blue background. Conserved bases are in upper case and start/stop codons are highlighted. **b**, Experimental design. The *A. nidulans* sequence from 26 nucleotides upstream of each uORF to four codons in the protein-coding gene was fused with a firefly luciferase gene. Controls were generated with start codons for both the uORF and the luciferase gene (+/+), the luciferase gene only (-/+), and the uORF only (+/-). Starts were deleted by alteration to ATT. mRNA from each construct was used to programme a cell-free translation system. **c**, Results of translation assays. The luciferase activity of all constructs (normalized to the -/+ construct) is shown on the y axis. Error bars show the average of the absolute deviation from the mean. The autoradiogram shows ³⁵S-Met-labelled firefly luciferase obtained by *in vitro* translation of the same mRNAs.

assembly of *A. oryzae* consists of transposable elements, compared to 3% of *A. nidulans* and *A. fumigatus*. All three genomes contain essentially all major classes of eukaryotic transposable elements, although the overall variety is relatively low. A number of unusual features were also observed in *Aspergillus* transposable elements (Supplementary Information). Copy numbers per family range from 1 to ~100; there are also many fragments and 'footprints' (evidence of recombinational excision). In addition, some members of all transposable element families longer than ~400 nucleotides are characterized by numerous C to T transitions. In *A. nidulans* and *A. fumigatus*, these predominantly affect cytosines in CpG and CpA doublets (no preference is apparent in *A. oryzae*). Moreover, repeat density is correlated with A + T richness in all three species (Fig. 2). The predominance of transition mutations is consistent with the operation of RIP (repeat-induced point mutation)⁴⁹, and all three aspergilli have a single predicted homologue (called DmtA) to the DNA methyltransferase *rid* that is essential for RIP in *N. crassa*⁴⁹. Apart from the putative *rid* homologue, no additional DNA methyltransferase genes were identified, consistent with failures to demonstrate methylation in these fungi. Although RIP has not been demonstrated in any *Aspergillus* species, if active it may be more similar to the mild form in *M. grisea*⁴⁹, as many transposable elements in these species are mutation-free.

Conclusion and perspective

The *A. nidulans* genome sequence and our comparative analysis with the genome sequences of *A. fumigatus* and *A. oryzae* have shed new light on the physiology of these fungi, as well as insight into aspects of genome evolution and gene regulation likely to be common to all eukaryotes. These results represent the initial step in realizing the full potential of these genomes. As a result of the genome analysis, efforts are underway to cross different isolates of *A. fumigatus* and *A. oryzae*. The identified conserved sequences also represent a rich set of targets for further experimental investigation. These efforts and ongoing sequencing projects for additional aspergilli promise to change fundamentally our understanding of this important group of medically, industrially and scientifically relevant fungi.

METHODS

Complete details of the methods used are available in Supplementary Information.

***A. nidulans* sequencing, assembly and analysis.** The *A. nidulans* genome strain FGSC A4 was sequenced by the WGS method to a depth of 10X. An additional 3X sequence coverage was provided by Monsanto (<http://www.monsanto.com/>). All sequence was assembled using Arachne. The *A. nidulans* genome was annotated as described in Supplementary Information. *A. fumigatus* and *A. oryzae* were assembled and annotated as described separately^{2,3}.

Phylogenetic analysis. A total of 3,034 predicted orthologues among *A. nidulans*, *A. oryzae*, *A. fumigatus*, *N. crassa* and *F. graminearum* were aligned at the protein level, back translated to DNA codons, and large gaps (>9 bp) were removed. Random sets of 20 DNA alignments were concatenated and passed to Phylip to generate 100 bootstrap replicates and a consensus maximum parsimony tree. Maximum likelihood trees were calculated on each replicate and a consensus tree was produced. Repeating with 1,000 bootstrap samples led to essentially identical results. For rooting with *C. immitis*, *C. immitis* orthologous CDS regions based on TblastN were retrieved, translated and aligned at the protein level with the aligned portions of the *Aspergillus* genes. Maximum parsimony trees were then generated and filtered for those with 100% bootstrap values at all nodes. The *C. immitis* sequence is available at http://www.broad.mit.edu/annotation/fungi/coccidioides_immitis/.

Hierarchical synteny mapping and branch-specific rearrangements. Protein homology anchors were detected using BlastP and filtered to retain only hits scoring >80% of the score of the best hit to each query protein. Contiguous sets of homologous proteins with conserved order and orientation were grouped into clusters. Pairs of clusters were then merged into successively larger clusters by tolerating successively larger breaks between clusters. Branch-specific breaks were determined by identifying regions without breaks between a reference species and query species, and then identifying breaks in that region between the reference and the third (target) species. Such breaks were considered specific to

the third species. Breaks were classified according to the pattern of apparent rearrangement.

Identification of non-coding conserved sequences. Genomic sequence for predicted three-way orthologues, including 1 kb upstream and downstream, were multiply aligned using Mlagan³³. An additive scoring function was used to identify maximal scoring subsequences. A cutoff of 22 was used to define HCSs unlikely to occur by chance ($P < 0.015$) according to models of neutral sequence and random sequence alignment. To model alignments of neutrally evolving sequence, aligned columns of fourfold degenerate sites were selected randomly and concatenated. To model alignments of random sequence, randomly selected intergenic regions from each genome were aligned. Fixed and variable length alignments were generated for both models. For each model, 1,000 simulated alignments were generated and maximal scoring subsequences were identified. The number of subsequences for each score was normalized by the number of aligned nucleotides. These rates were used to determine the score cutoff above.

Prediction of functional motifs. Each HCS was represented as a position-specific probability matrix (PSPM) derived from the three-way alignment. Each PSPM was compared to each other PSPM and matching PSPMs were clustered. Local multiple alignments for each cluster were generated and the resulting multiple alignments and corresponding weight matrices were termed conpats. For each conpat, we used the corresponding weight matrix with MAST to identify instances where the conpat co-occurred upstream or downstream of orthologous genes in all three of the aspergilli. A series of conservation tests was then applied to the set of predicted co-occurrences for each conpat as described in Supplementary Information.

Prediction and validation of upstream open reading frames. Genome sequences for three-way orthologues, including 1,000 bp upstream and downstream, were multiply aligned using Mlagan. For 25% of *A. nidulans* genes 5' UTR sequences were predicted from EST alignments. On the basis of the length distribution of these EST-predicted UTRs, we used 60 bp upstream of predicted AUG codons as a conservative estimate of 5' UTRs for genes lacking ESTs. When no ESTs were available, UTRs for orthologues where considered when all three annotated start codons aligned within 40 bp. We identified uORFs ≥ 12 bp, with a maximum 1-bp overlap with the protein-coding gene's ATG. Conserved uORFs were identified as those for which the start and stop codons were exactly aligned within the multiple alignments. To experimentally validate uORFs, synthetic oligonucleotides containing each uORF, 26 nucleotides upstream of the uORF, the region between the uORF and the protein coding AUG, and the first four codons of the protein coding gene were fused in frame with a firefly luciferase gene. Three different control constructs were also generated. Capped and polyadenylated synthetic mRNA were prepared from each construct and equal amounts were used to programme cell-free extracts from *N. crassa*. Differential translation between the intact construct and the controls was measured using a luciferase activity assay as well as through the production of ³⁵S-Met-labelled firefly luciferase obtained by *in vitro* translation of the same mRNAs.

The *A. nidulans* genome sequence is available at <http://www.broad.mit.edu/> and has been deposited at DDBJ/EMBL/GenBank under the project accession AACD00000000.

Received 31 May; accepted 19 October 2005.

1. Timberlake, W. E. & Marshall, M. A. Genetic engineering of filamentous fungi. *Science* **244**, 1313–1317 (1989).
2. Nierman, W. *et al.* Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* doi:10.1038/nature04332 (this issue).
3. Machida, M. *et al.* Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* doi:10.1038/nature04300 (this issue).
4. Pontecorvo, G. & Kafer, E. Genetic analysis based on mitotic recombination. *Adv. Genet.* **9**, 71–104 (1958).
5. Pontecorvo, G., Roper, J. A., Hemmons, L. M., Macdonald, K. D. & Bufton, A. W. The genetics of *Aspergillus nidulans*. *Adv. Genet.* **5**, 141–238 (1953).
6. Peterson, S. W. in *Integration of Modern Taxonomic Methods for Penicillium and Aspergillus Classification* (eds Samson, R. A. & Pitt, J. I.) 323–355 (Harwood Academic Publishers, Amsterdam, The Netherlands, 2000).
7. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003).
8. Verweij, P. E. *et al.* Phylogenetic relationships of five species of *Aspergillus* and related taxa as deduced by comparison of sequences of small subunit ribosomal RNA. *J. Med. Vet. Mycol.* **33**, 185–190 (1995).
9. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
10. Eichler, E. E. & Sankoff, D. Structural dynamics of eukaryotic chromosome evolution. *Science* **301**, 793–797 (2003).
11. Coghlan, A. & Wolfe, K. H. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.* **12**, 857–867 (2002).

12. Burt, D. W. *et al.* The dynamics of chromosome evolution in birds and mammals. *Nature* **402**, 411–413 (1999).
13. Sharakhov, I. V. *et al.* Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* **298**, 182–185 (2002).
14. Slamovits, C. H., Fast, N. M., Law, J. S. & Keeling, P. J. Genome compaction and stability in microsporidian intracellular parasites. *Curr. Biol.* **14**, 891–896 (2004).
15. Paoletti, M. *et al.* Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Curr. Biol.* **15**, 1242–1248 (2005).
16. Metznerberg, R. L. & Glass, N. L. Mating type and mating strategies in *Neurospora*. *Bioessays* **12**, 53–59 (1990).
17. Dyer, P. S. in *Proc. 7th Int. Mycol. Congr. Abstr.* 1095, 331 (International Mycological Congress, Oslo, 2002).
18. Dyer, P. S., Paoletti, M. & Archer, D. B. Genomics reveals sexual secrets of *Aspergillus*. *Microbiology* **149**, 2301–2303 (2003).
19. Yun, S. H., Berbee, M. L., Yoder, O. C. & Turgeon, B. G. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proc. Natl Acad. Sci. USA* **96**, 5592–5597 (1999).
20. Poggeler, S. Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. *Curr. Genet.* **42**, 153–160 (2002).
21. Butler, G. *et al.* Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc. Natl Acad. Sci. USA* **101**, 1632–1637 (2004).
22. Geiser, D. M., Timberlake, W. E. & Arnold, M. L. Loss of meiosis in *Aspergillus*. *Mol. Biol. Evol.* **13**, 809–817 (1996).
23. Geiser, D. M., Frisvad, J. & Taylor, J. Evolutionary relationships in *Aspergillus* section *Fumigati* inferred from partial β -tubulin and hydrophobin DNA sequences. *Mycologia* **90**, 831–845 (1998).
24. Perkins, D. D. Mating-type switching in filamentous ascomycetes. *Genetics* **115**, 215–216 (1987).
25. Nauta, M. J. & Hoekstra, R. F. Evolution of reproductive systems in filamentous ascomycetes. II. Evolution of hermaphroditism and other reproductive strategies. *Heredity* **68**, 537–546 (1992).
26. Coppin, E., Debuchy, R., Arnaise, S. & Picard, M. Mating types and sexual development in filamentous ascomycetes. *Microbiol. Mol. Biol. Rev.* **61**, 411–428 (1997).
27. Lucas, J. A., Dyer, P. S. & Murray, T. D. Pathogenicity, host specificity, and population biology of *Tapesia* spp. causal agents of eyespot disease of cereals. *Adv. Bot. Res.* **33**, 225–258 (2000).
28. Kang, S., Chumley, F. G. & Valent, B. Isolation of the mating-type genes of the phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. *Genetics* **138**, 289–296 (1994).
29. Kerenyi, Z., Moretti, A., Waalwijk, C., Olah, B. & Hornok, L. Mating type sequences in asexually reproducing fusarium species. *Appl. Environ. Microbiol.* **70**, 4419–4423 (2004).
30. Magee, B. B. & Magee, P. T. Induction of mating in *Candida albicans* by construction of MTL α and MTL β strains. *Science* **289**, 310–313 (2000).
31. Hull, C. M., Raisner, R. M. & Johnson, A. D. Evidence for mating of the “asexual” yeast *Candida albicans* in a mammalian host. *Science* **289**, 307–310 (2000).
32. Sharon, A. *et al.* An asexual fungus has the potential for sexual development. *Mol. Gen. Genet.* **251**, 60–68 (1996).
33. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).
34. Morozov, I. Y., Martinez, M. G., Jones, M. G. & Caddick, M. X. A defined sequence within the 3' UTR of the *areA* transcript is sufficient to mediate nitrogen metabolite signalling via accelerated deadenylation. *Mol. Microbiol.* **37**, 1248–1257 (2000).
35. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
36. Hoffmann, B., Valerius, O., Andermann, M. & Braus, G. H. Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene *cpcA* of filamentous fungus *Aspergillus nidulans*. *Mol. Biol. Cell* **12**, 2846–2857 (2001).
37. Natarajan, K. *et al.* Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol. Cell. Biol.* **21**, 4347–4368 (2001).
38. Eckert, S. E., Kubler, E., Hoffmann, B. & Braus, G. H. The tryptophan synthase-encoding *trpB* gene of *Aspergillus nidulans* is regulated by the cross-pathway control system. *Mol. Gen. Genet.* **263**, 867–876 (2000).
39. Valerius, O. *et al.* Regulation of *hisH* transcription of *Aspergillus nidulans* by adenine and amino acid limitation. *Fungal Genet. Biol.* **32**, 21–31 (2001).
40. Gerber, A. P., Herschlag, D. & Brown, P. O. Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* **2**, E79 (2004).
41. Geballe, A. P. & Sachs, M. S. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 595–614 (Cold Spring Harbor Laboratory Press, New York, 2000).
42. McCarthy, J. E. Posttranscriptional control of gene expression in yeast. *Microbiol. Mol. Biol. Rev.* **62**, 1492–1553 (1998).
43. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol.* **3**, REVIEWS0004 (2002).
44. Maggio-Hall, L. A. & Keller, N. P. Mitochondrial β -oxidation in *Aspergillus nidulans*. *Mol. Microbiol.* **54**, 1173–1185 (2004).
45. Hiltunen, J. K. *et al.* The biochemistry of peroxisomal β -oxidation in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **27**, 35–64 (2003).
46. Kikuchi, M. *et al.* Proteomic analysis of rat liver peroxisome: presence of peroxisome-specific isozyme of Lon protease. *J. Biol. Chem.* **279**, 421–428 (2004).
47. Balzi, E., Wang, M., Leterme, S., Van Dyck, L. & Goffeau, A. PDR5, a novel yeast multidrug resistance conferring transporter controlled by the transcription regulator PDR1. *J. Biol. Chem.* **269**, 2206–2214 (1994).
48. Harris, S. D. & Momany, M. Polarity in filamentous fungi: moving beyond the yeast paradigm. *Fungal Genet. Biol.* **41**, 391–400 (2004).
49. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**, 417–423 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank M. Zody, X. Xie, M. Kellis, M. Kamal, J. Taylor, G. Turgeon and E. Lander for numerous helpful conversations, comments and critical readings of the paper. We also thank all members of the sequencing platform at the Broad Institute. We thank R. Dean for providing the BAC library used in sequencing. We especially thank M. Brudno for his help in using Mlagan to align the three genomes, and R. Morris for his many contributions to the *A. nidulans* sequencing project and subsequent analyses. This work was supported in part by grants from the NIH, as well as the NIH Research Supplement for Underrepresented Minorities, the DFG, the CMPB and the BBSRC.

Author Contributions B.W.B. developed and led the *A. nidulans* sequencing project. J.B. performed the assembly of *A. nidulans*. C.C. and J.E.G. analysed the phylogenetic relationship of the three organisms. J.E.G. performed the analysis of conserved synteny and genome evolution. R.F., B.M. and P.D. performed the comparative analysis of sexual reproduction genes. J.E.G., M.P., P.D. and B.M. analysed the mating-type loci. S.E.C. and J.E.G. analysed conserved non-coding sequences and computationally analysed uORFs. S. Batzoglou and S.-I.L. assisted with the use of Mlagan. M.B., C.C.S. and M.S.S. performed the experimental uORF validation. M.C., M.H., G.H.B., O.D. and C.D. analysed transcription, known transcription factors and binding sites. A.P. and S.G.J. predicted non-coding RNAs using PFAM and RFAM. J.C., V.K., J.J., S.P. and J.E.G. analysed repeat sequences and RIP. M. Farman analysed telomeres and subtelomeric gene content. M.H. analysed the *Aspergillus* peroxisomes. S.H. and M. Momany analysed hyphal growth and RhoGTPases. J.R.W. and W.C.N. provided the sequence and annotation for *A. fumigatus*. T.T., T.K., K.A. and M. Machida provided the sequence and annotation for *A. oryzae*. D.W.D. co-ordinated interactions between the different sequencing centres and scientific communities. J.E.G. and S.A.O. coordinated the comparative analyses. J.E.G. wrote and edited the paper, and produced the figures.

Author Information The *A. nidulans* genome sequence has been deposited at DDBJ/EMBL/GenBank under the project accession AACD00000000. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.E.G. (jgalag@mit.edu).

ARTICLES

Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation

Wolfgang Fischle¹, Boo Shan Tseng², Holger L. Dormann¹, Beatrix M. Ueberheide³, Benjamin A. Garcia³, Jeffrey Shabanowitz³, Donald F. Hunt^{3,4}, Hironori Funabiki² & C. David Allis¹

Tri-methylation of histone H3 lysine 9 is important for recruiting heterochromatin protein 1 (HP1) to discrete regions of the genome, thereby regulating gene expression, chromatin packaging and heterochromatin formation. Here we show that HP1 α , β , and γ are released from chromatin during the M phase of the cell cycle, even though tri-methylation levels of histone H3 lysine 9 remain unchanged. However, the additional, transient modification of histone H3 by phosphorylation of serine 10 next to the more stable methyl-lysine 9 mark is sufficient to eject HP1 proteins from their binding sites. Inhibition or depletion of the mitotic kinase Aurora B, which phosphorylates serine 10 on histone H3, causes retention of HP1 proteins on mitotic chromosomes, suggesting that H3 serine 10 phosphorylation is necessary for the dissociation of HP1 from chromatin in M phase. These findings establish a regulatory mechanism of protein-protein interactions, through a combinatorial readout of two adjacent post-translational modifications: a stable methylation and a dynamic phosphorylation mark.

Genomic DNA within the eukaryotic nucleus is organized into distinct chromosomal domains^{1,2}. Structural and functional changes in the organization and dynamics of such specialized chromatin areas are key to controlling genome function³. Cytologically defined domains of euchromatin and heterochromatin have been functionally linked to gene content and activity. Euchromatin generally keeps genes competent for transcription, whereas heterochromatin contains predominantly transcriptionally silent genes and includes specialized chromosome structures such as centromeres and telomeres^{4,5}.

Members of the heterochromatin protein 1 (HP1) family have important roles in heterochromatin organization^{4,6}. The three isoforms of HP1 (α , β , and γ) in higher eukaryotes have been associated with constitutive (that is, pericentric and telomeric) heterochromatin and some forms of facultative (that is, developmentally regulated) heterochromatin⁵. Although all HP1 isoforms are localized predominantly to pericentric heterochromatin, HP1 β and γ can also be found at euchromatic sites⁷, where they are presumably involved in gene repression^{5,6}. During mitosis, a fraction of HP1 α stays associated with (peri-)centromeric chromosome regions^{7–9}, and the single HP1 homologue in *Schizosaccharomyces pombe*, Swi6, is required for proper chromosome segregation and cohesion of sister centromeres¹⁰.

Recruitment of HP1 proteins to certain sites of the genome involves interactions with multiple components of chromatin. In particular, the methylation of histone H3 on lysine 9 (H3K9me) is important for bringing HP1 to distinct chromosomal areas^{11–13}, and both tri-methylated H3K9 (H3K9me3) and HP1 are thought to be crucial for establishing and maintaining domains of heterochromatin^{14,15}. Indeed, an amino-terminal chromodomain found in all HP1 proteins specifically interacts with H3K9me *in vitro*^{16–18}, particularly in its di- and tri-methylated states^{19,20}. Although binding of HP1 to

H3K9me is fairly weak, the overlap of HP1 proteins and H3K9me3 at heterochromatic sites of different cell types validates the biological importance of this effector–mark interaction. Such an apparent increase in cellular binding strength may be due to the cooperative effects of HP1 dimerization and additional stabilizing interactions with other chromatin factors^{5,11,13,21}.

In agreement with the rather weak HP1–H3K9me3 interaction, a large fraction of the cellular HP1 molecules are not stably incorporated into heterochromatin, but instead show rapid on/off kinetics from their subnuclear target areas^{9,22,23}. The dynamic properties of HP1 and other components of chromatin domains may be critical for creating an ever-changing—but overall steady—architectural framework within which nuclear processes can take place²¹. The mechanisms that control the dynamic interaction between HP1 and H3K9me3, however, have not been established.

Methylation-phosphorylation of histone H3 in M phase

As the bulk of mammalian HP1 proteins dissociate from chromatin during mitosis^{7–9}, we sought to investigate the role of the HP1–H3K9me3 interaction in the dynamic association of HP1 proteins with chromatin in the context of the cell cycle. As observed for the different HP1 isoforms in mammalian cell lines^{7,12,18,24,25}, we found that HP1 β localized to dot-like structures in the cell nucleus of 10T1/2 cells at interphase (Fig. 1a). These dot-like structures contain constitutive pericentric heterochromatin and show strong staining with DAPI²¹. Consistent with the involvement of H3K9 methylation in HP1 recruitment, the H3K9me3 mark was found enriched in the same subnuclear areas in interphase cells. In contrast, this colocalization was not observed in M-phase cells, in which HP1 β was diffusely distributed throughout the cell^{7–9} while the H3K9me3 mark remained tightly localized to condensed chromosomes. Western blot (Fig. 1b) and quantitative mass spectrometric (Fig. 1d) analyses

¹Laboratory of Chromatin Biology and ²Laboratory of Chromosome and Cell Biology, The Rockefeller University, New York, New York 10021, USA. ³Departments of Chemistry and ⁴Pathology, University of Virginia, Charlottesville, Virginia 22904, USA.

of histones isolated from cells arrested at M phase by nocodazole revealed that the overall level of H3K9me3 is not altered when cells enter mitosis. Similarly, the cellular protein levels of the different HP1 isoforms remained unchanged (Fig. 1b). These analyses indicate that HP1 dissociates from mitotic chromatin without loss of the 'recruiting' H3K9me3 mark, and suggest that the HP1–H3K9me3

interaction must be somehow interrupted as cells enter M phase.

We next sought to analyse whether additional modifications occur in the vicinity of the K9me3 mark on the H3-tail during mitosis. We therefore purified H3 from HeLa cells arrested at M phase and used mass spectrometry to examine the modification pattern on the H3-tail^{12,26}. This approach revealed the existence of a phospho-mark (ph)

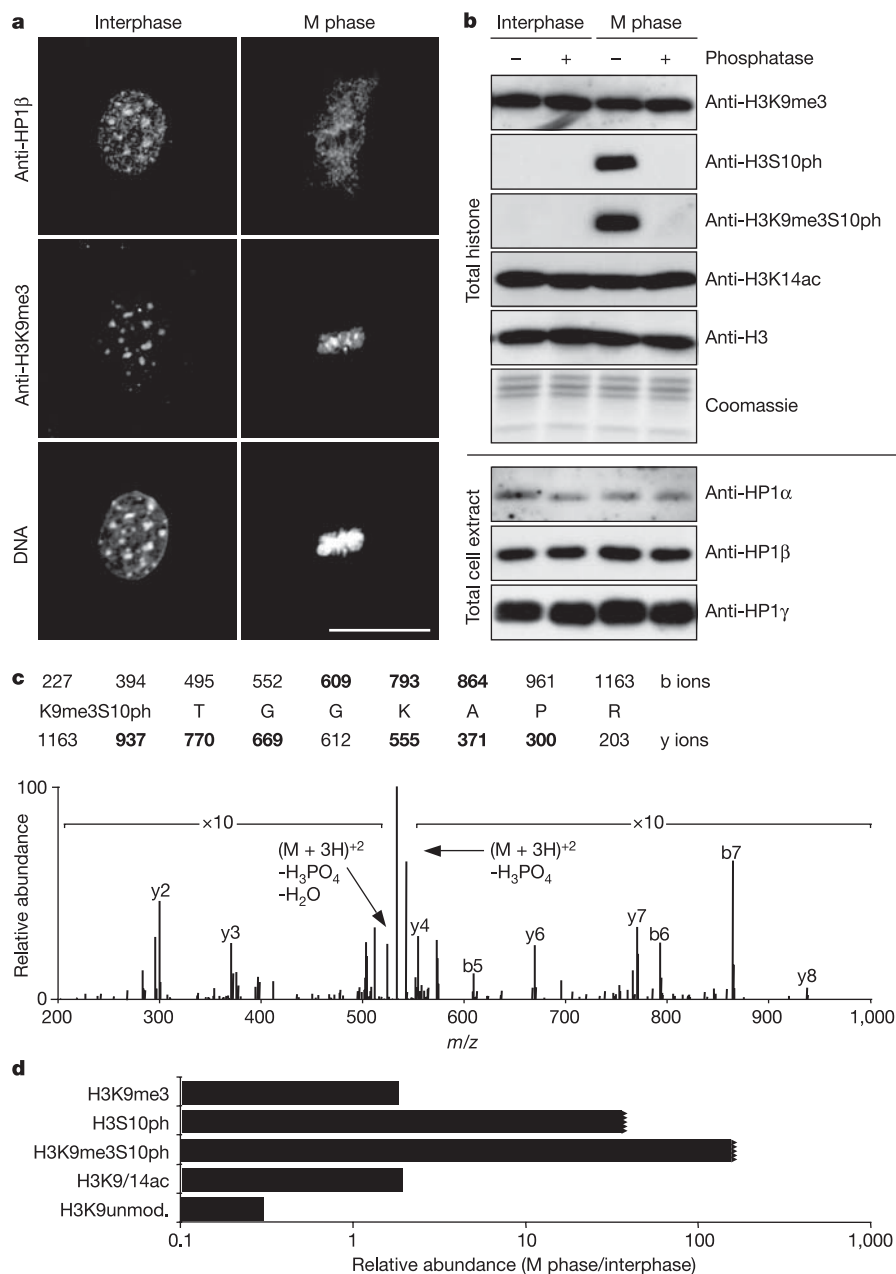


Figure 1 | Coordinate behaviour of HP1 and mitotic phosphorylation of H3S10 in the context of H3K9me3. **a**, Distribution of HP1 β in10T1/2 fibroblast cells during interphase (left) and mitosis (right, metaphase stage shown), analysed by immunostaining. DNA was stained with DAPI. Scale bar, 10 μ m. **b**, Western blot analysis of histones (top) and cell extracts (bottom) prepared from asynchronously growing (interphase) or nocodazole-arrested (M phase) HeLa cells after treatment with or without alkaline phosphatase. Gels stained with Coomassie were used as loading controls. See Supplementary Fig. S1 for characterization of the H3K9me3S10ph-specific antibody. **c**, Tandem mass spectrometry (MS/MS) spectrum recorded on doubly protonated ($M + 2H$) $^{2+}$ ions derived from residues 9–17 of histone H3 isolated from mitotically arrested HeLa cells. This spectrum confirms the presence of K9me3 and S10ph on the same H3-tail. The measured (582.3127 Da) and calculated (582.3115 Da) weights

for the parent $(M+2H)^{2+}$ ions are within 2.2 p.p.m. Predicted masses for singly charged fragment ions of type b (acylium ions containing the N terminus) and type y (truncated peptides containing the C terminus) are shown above and below the peptide sequence; those observed in the spectra are bold. Abundances of fragments other than those corresponding to $(M+2H)^{2+}-H_3P0_4$ and $(M+2H)^{2+}-(H_3P0_4 \text{ and } H_2O)$ are amplified by a factor of 10. **d**, Relative abundance of H3 N-terminal peptides isolated from asynchronous (interphase) or nocodazole-arrested (M phase) HeLa cells as determined by mass spectrometry¹². Acetylation marks on H3K9 and H3K14 cannot be discriminated by this method (H3K9/K14ac). H3unmod, unmodified H3-tail. Note that neither H3S10ph nor H3K9me3S10ph was detected in the asynchronous sample. Plotted minimal values for these entries used the detection limit observed in the analysis of the asynchronous sample ($\sim 0.01\%$).

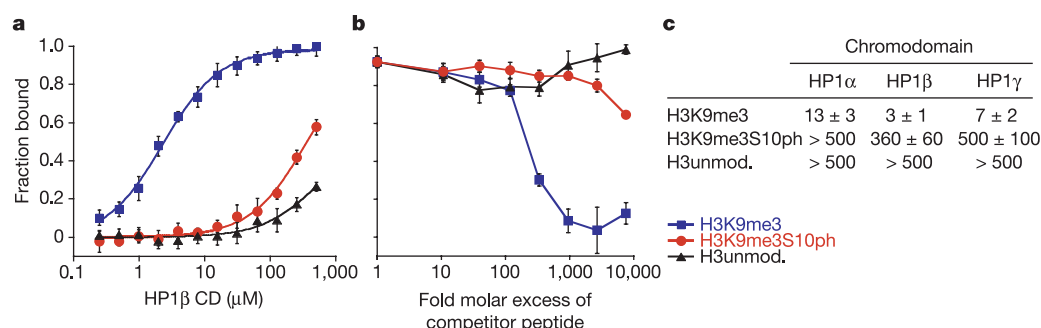


Figure 2 | Binding of HP1 to an H3K9me3 peptide is impaired by phosphorylation of H3S10. **a**, Binding of the HP1β chromodomain (CD) to the indicated H3-tail peptides was analysed by fluorescence polarization assays. Data show average ± s.d. for three independent experiments. **b**, Competition experiment between a fluorescein-labelled H3K9me3 peptide bound to the HP1β chromodomain (95% saturation) and the

indicated unlabelled H3 peptides, analysed by fluorescence polarization. Data show average ± s.d. for three independent experiments. **c**, Dissociation constants (K_d in μM) for the interaction of the chromodomains of the different human HP1 isoforms with H3K9me and H3K9meS10ph peptides. Values represent averages ± s.d. from ≥ 3 independent experiments.

on Ser 10 (H3S10) next to K9me3 (Fig. 1c). Dual-mark combinations of H3K9me1S10ph and H3K9me2S10ph were also detected (data not shown). No phosphorylation on Ser 10, alone or in combination with K9me, could be found on H3 purified from cells at interphase. To further validate the existence of H3K9me3S10ph, we raised an antiserum that specifically recognizes this dual-mark combination (Supplementary Fig. S1). Western blot analysis of histones prepared from nocodazole-arrested cells indeed verified the occurrence of H3K9me3S10ph specifically in M phase chromatin (Fig. 1b). These data are consistent with H3S10ph being a 'mitotic mark' that first appears at pericentromeric heterochromatin in late G2 (refs 27, 28), and raise the possibility that the dual-mark combination of H3K9me3S10ph controls HP1–chromatin binding.

H3S10ph inhibits the HP1–H3K9me3 interaction

To investigate the effect of H3S10 phosphorylation on the HP1–H3K9me interaction, we analysed the binding of HP1 to dually modified H3K9me3S10ph peptides using fluorescence polarization measurements. In direct binding (Fig. 2a) and indirect competition experiments (Fig. 2b), we detected a significant loss in the affinity of HP1 chromodomains for a dually modified H3K9me3S10ph peptide, compared with the interaction between HP1 and a H3K9me3 peptide. Although the HP1 chromodomains reproducibly interacted better with the H3K9me3S10ph peptide than with the unmodified (unmethylated) control H2 peptide, we note that the affinity for the H3K9me3 mark was two orders of magnitude lower for all HP1 isoforms when the H3S10ph modification was present (Fig. 2c). These effects were not restricted to HP1 chromodomains, as experiments using full-length recombinant proteins showed the same loss of binding (Supplementary Fig. S2a). Mutating the conserved glutamic acid residue within the H3-binding groove of HP1 that forms a hydrogen bond with H3S10 (ref. 19) confirmed the importance of this interaction for HP1–H3K9me binding (Supplementary Fig. S2). Our biophysical studies reveal a putative function for the dual-mark combination of H3K9me3S10ph in the inhibition of HP1 recruitment. The competition studies further suggest that binding of HP1 to the methylated H3-tail is fully reversible and highly dynamic, thereby supporting the rapid exchange of HP1 from heterochromatin^{9,22,23}.

Phosphorylation of H3S10 removes HP1 from H3K9me3

Although we detected dually modified H3K9me3S10ph in mitotic cells (Fig. 1), earlier studies have reported a reduced *in vitro* activity of the principal mitotic H3S10 kinase, Aurora B (ref. 29), on K9-methylated substrates compared with the unmodified H3-tail³⁰. As Aurora B is activated by additional factors within the chromosomal passenger complex (CPC)^{31–33}, we analysed the native H3S10 kinase

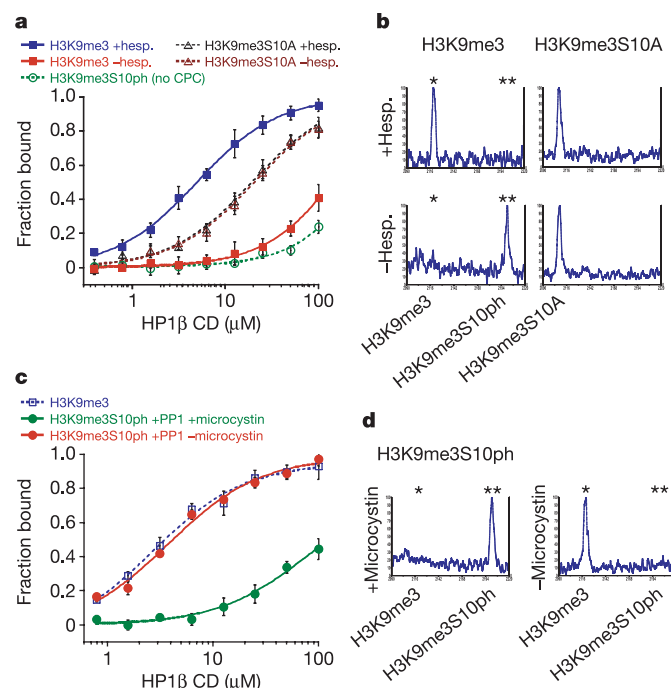


Figure 3 | Reversible phosphorylation of H3S10 disrupts the HP1–H3K9me3 interaction. **a**, Loss of HP1 binding to H3K9me3 by phosphorylation of H3S10. Interaction of the HP1β chromodomain (CD) with H3K9me3 and H3K9me3S10A peptides after phosphorylation by CPC in the presence (+hesp.) or absence (–hesp.) of hesperadin, measured by fluorescence polarization ($n = 3$; error bars show s.d.). **b**, Substrate histone peptides from the phosphorylation reactions in **a** containing 50 μM HP1β chromodomain were analysed by MALDI-TOF mass spectrometry. Left panels show loss of H3K9me3 signal (*) over the course of the reaction, and appearance of H3K9me3S10ph peptide (**). Right panels show control H3K9me3S10A reaction. **c**, Gain of HP1 binding to a H3K9me3S10ph peptide after dephosphorylation of H3S10. Interaction of the HP1β chromodomain with a fluorescein-labelled H3K9me3S10ph peptide after dephosphorylation by phosphatase PP1 in the presence or absence of microcystin LR, measured by fluorescence polarization ($n = 3$; error bars show s.d.). **d**, Substrate histone peptides from the dephosphorylation reactions in **c** containing 50 μM HP1β chromodomain were analysed by MALDI-TOF mass spectrometry. Loss of H3K9me3S10ph signal (**) over the course of the reaction, and appearance of the H3K9me3 peptide (*) are shown.

complex on modified H3-tail peptides. The relative ability of Aurora B to phosphorylate H3S10 when K9 is methylated to various degrees is very similar to that measured with the unmodified substrate (Supplementary Fig. S3).

We then tested whether enzymatic phosphorylation of H3S10 could eject HP1 bound to H3K9me3. Preliminary experiments using the CPC kinase complex indicated that an H3K9me3 peptide could be phosphorylated even in the presence of a large molar excess of HP1 chromodomain (Supplementary Fig. S4). Interaction of a fluorescein-labelled H3K9me3 peptide with increasing amounts of HP1 β chromodomain after CPC kinase reaction was recorded using fluorescence polarization measurements. Similar to the reduced interaction between HP1 and the dually modified H3K9me3S10ph peptide (Fig. 2), we detected a significant loss of HP1 β chromodomain binding to the methylated H3 peptide after the phosphorylation reaction. However, this change in binding was not observed when Aurora B was inhibited by the small molecule hesperadin^{34,35} (Fig. 3a) or when a H3K9me3 peptide containing a Ser 10 to Ala 10 substitution (S10A, which shows about fivefold reduced interaction with the HP1 chromodomain) was used¹⁹. Mass spectrometric analysis of the H3-tails after the kinase reaction verified phosphorylation of the H3K9me3 peptide, but not the control H3K9me3S10A peptide, confirming the specificity of CPC-mediated phosphorylation at the H3S10 site (Fig. 3b). Notably, the HP1 β chromodomain was neither phosphorylated nor degraded over the course of the reaction, and similar results were obtained with the chromodomains of HP1 α and HP1 γ as well as full-length HP1 β protein (Supplementary Fig. S4). In a reverse reaction sequence, we detected a gain-of-binding of the HP1 β chromodomain to an H3K9me3S10ph peptide in the presence of protein phosphatase 1 (PP1), an enzyme known to dephosphorylate H3S10ph (ref. 28) (Fig. 3c, d). This gain in binding was sensitive to the phosphatase inhibitor microcystin LR. On the basis of these observations, we reason that transient (reversible) phosphorylation of H3S10 during mitosis might control the dynamic interaction between HP1 and H3K9me3.

H3K9me3S10ph concurs with mitotic HP1 relocation

Next, we examined the effect of H3S10 phosphorylation on the association of HP1 with chromatin *in vivo* and investigated the role of dually modified H3K9me3S10ph in the release of HP1 from chromatin at the onset of mitosis. We stained asynchronously growing

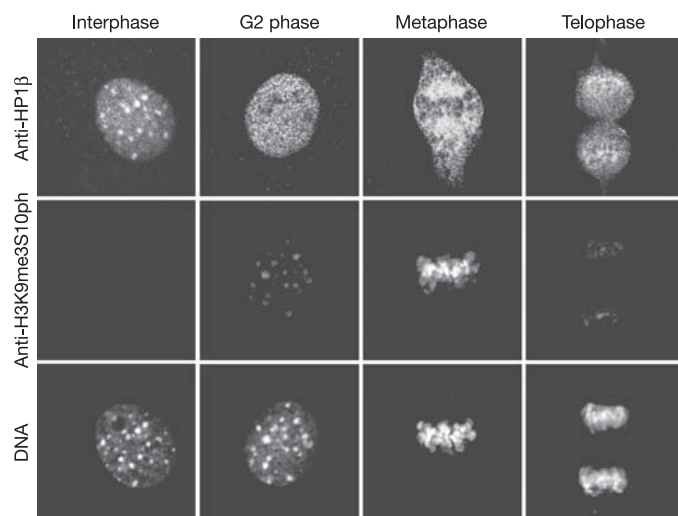


Figure 4 | Temporal occurrence of the dual K9me3S10ph epitope on H3 coincides with dissociation of HP1 from mitotic chromatin. 10T1/2 cells were stained with the indicated primary antibodies. DNA was stained with DAPI. Columns represent cells at the indicated stages of the cell cycle. Scale bar, 10 μ m.

cultures of 10T1/2 cells with H3K9me3S10ph-specific antiserum and analysed when this dual-mark occurred during the cell cycle (Fig. 4). As expected from our western blot and mass spectrometric analyses of H3 (Fig. 1), the dually modified H3 epitope was not observed in interphase cells, which showed a dot-like subnuclear distribution of HP1 β (Fig. 4, interphase). H3K9me3S10ph was only detected in cells that had lost this characteristic HP1 β localization pattern and showed diffuse HP1 β nuclear staining. There, the dual-mark combination was restricted to the dot-like areas showing strong staining with DAPI (Fig. 4, G2). Upon chromosome condensation (Fig. 4, metaphase), HP1 β was found diffusely distributed throughout the cell, and little overlap with the anti-H3K9me3S10ph and DAPI staining was detected. Only as cells exited mitosis and phosphorylation of H3S10 disappeared did we observe re-association of HP1 β with chromatin. Similar observations were made in HeLa cells and for the α and γ isoforms of HP1 (data not shown). In agreement with the dot-like appearance of the anti-H3K9me3S10ph immunostaining at the onset of mitosis, we detected this dual-mark combination enriched at the centromeric and pericentromeric regions with a more spotted appearance on the chromosome arms of metaphase

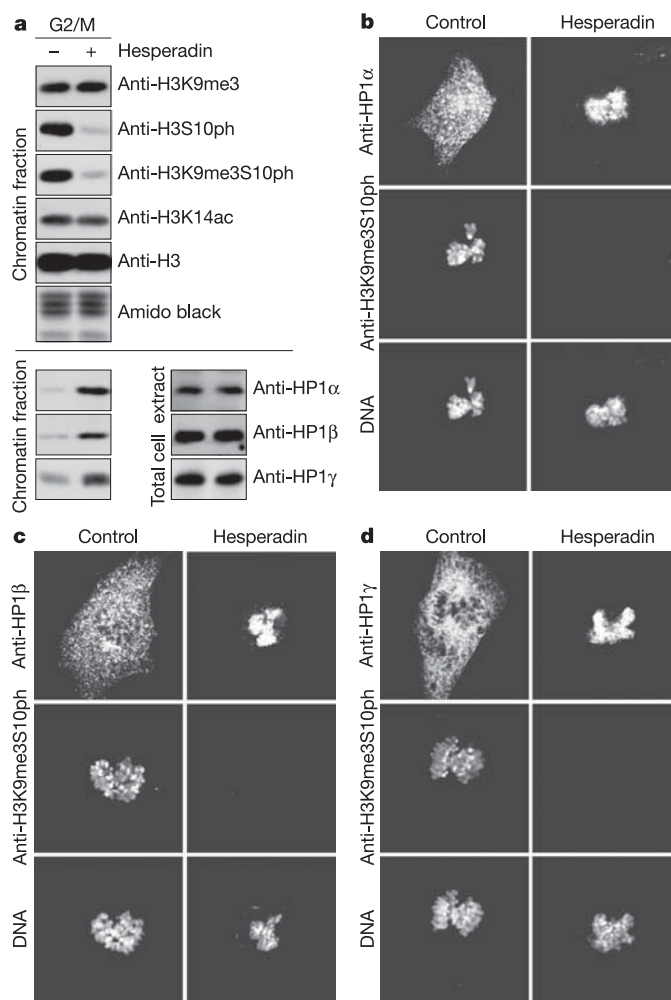


Figure 5 | Inhibition of Aurora B kinase results in retention of HP1 on M-phase chromatin. **a**, HeLa cells synchronized at the G2/M transition were treated with 200 nM of the Aurora B inhibitor hesperadin (+) or with vehicle (–). Total cellular chromatin or whole-cell extracts were analysed by western blot. Blots stained with amido black were used as loading controls. **b–d**, Immunofluorescence staining of 10T1/2 cells at M phase, with or without treatment with 200 nM hesperadin, using anti-HP1 α (**b**), anti-HP1 β (**c**) and anti-HP1 γ (**d**) primary antibodies. DNA was stained with DAPI. Scale bars, 10 μ m.

chromosome spreads (Supplementary Fig. S5). Together, our immunofluorescence analyses link the dissociation of HP1 from condensed chromatin in M phase to the temporal and local occurrence of the dual-mark combination of H3K9me3S10ph.

Lack of H3S10ph retains HP1 on mitotic chromosomes

To examine whether phosphorylation of H3S10 is causally linked to the release of HP1 from chromatin at the onset of mitosis, we analysed the mitotic behaviour of HP1 when Aurora B is inhibited. Treatment of G2/M-synchronized HeLa cells with hesperadin resulted in the loss of the mitotic H3S10ph mark³⁵ and, as predicted, the elimination of the dual H3K9me3S10ph epitope (Fig. 5a). In contrast, no change in the global levels of H3K9me3 and acetylated Lys 14 on H3 (H3K14ac) were observed, and the overall protein levels of HP1 α , β and γ were not affected. Similar effects of hesperadin on the phosphorylation of H3S10 and the loss of the dual-mark combination of H3K9me3S10ph were observed in 10T1/2 cells (Supplementary Fig. S6).

We then analysed the effect of inhibiting mitotic H3S10 phosphorylation on HP1 distribution and localization in M-phase cells. As observed in other cell types^{7–9}, HP1 α is partly retained on M-phase chromosomes and shows an otherwise diffuse distribution throughout 10T1/2 cells at stages when chromatin is condensed. In contrast, HP1 β and γ , are largely absent from compacted M-phase chromosomes in these cells (Fig. 5b–d, control). Inhibition of Aurora B kinase activity by hesperadin, as revealed by the absence of anti-H3K9me3S10ph immunostaining, resulted in dramatic changes in the localization of all HP1 isoforms in M-phase cells (Fig. 5b–d, hesperadin): HP1 α , β , and γ remained localized to condensed chromatin in the absence of mitotic H3S10 phosphorylation. Hesperadin treatment did not affect the distribution of HP1 in interphase cells (Supplementary Fig. S6). Similar results were obtained with ectopically expressed HP1–green fluorescent protein (GFP) fusion proteins in HEp-2 cells and after knockdown of Aurora B by RNA interference (RNAi) (Supplementary Figs S7, S8). Furthermore, biochemical fractionation of chromatin isolated

from hesperadin-treated HeLa cells verified the increased association of HP1 α , β and γ with mitotic chromatin when H3S10 phosphorylation is inhibited (Fig. 5a). Together, these experiments indicate that phosphorylation of H3S10 by Aurora B is necessary for the dissociation of HP1 from mitotic chromatin.

HP1 binds to M-phase chromosomes lacking H3S10ph

To further investigate the molecular mechanism by which HP1 dissociates from chromosomes in M phase and to avoid the experimental limitations of tissue culture systems, we turned to *Xenopus* egg extracts, from which metaphase chromosomes can be purified in a well-controlled manner³⁶. Consistent with the results in 10T1/2 and HeLa cells, we observed the dual-mark combination of H3K9me3S10ph (Supplementary Fig. S9) and the dissociation of *Xenopus* HP1 α (xHP1 α)–GFP from chromosomes upon entry into M phase (data not shown). To examine whether this dissociation is caused by phosphorylation of H3S10 by Aurora B, we monitored the chromosome-binding activity of xHP1 in egg extracts depleted of the Aurora B-containing CPC complex (Δ CPC)³⁶, in which chromosomal H3S10ph was greatly decreased but the level of H3K9me3 was not affected (Supplementary Fig. S10). Indeed, using anti-xHP1 α antibodies (Supplementary Fig. S11), we detected increased binding of endogenous xHP1 α to metaphase chromosomes assembled in Δ CPC extracts compared with control extracts (Fig. 6a). To quantify this response, metaphase chromosomes were purified from extracts incubated with [³⁵S]-labelled xHP1 α , β and γ . At least sevenfold more xHP1 protein was associated with metaphase chromosomes assembled in Δ CPC extracts compared with control extracts (Fig. 6b).

We next sought to verify that this aberrant association of HP1 with M-phase chromosomes in Δ CPC extracts is dependent on chromodomain–H3K9me3 interaction. First, we examined the effect caused by mutating one of the three aromatic residues essential in HP1 chromodomains for binding H3K9me *in vitro*^{19,37} (see Supplementary Fig. S2). In agreement with the importance of the HP1 chromodomain in the chromosome-binding activity of xHP1 α , the

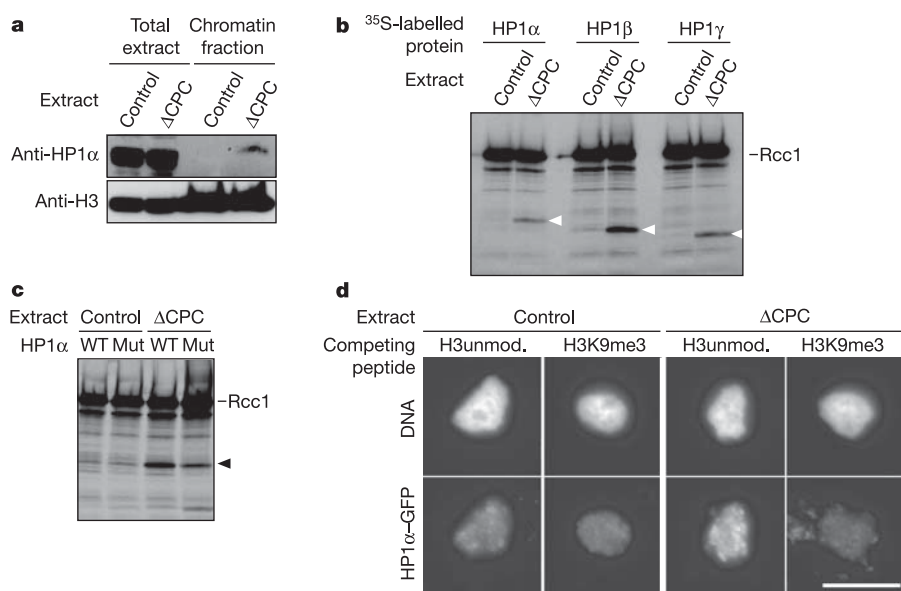


Figure 6 | HP1 binds more strongly to M-phase chromosomes in the absence of Aurora B. **a**, Western blot of control or Δ CPC *Xenopus* egg extracts (total extract) and purified metaphase chromosomes assembled in these extracts (chromatin fraction) using anti-xHP1 α and anti-H3 antibodies. **b**, Autoradiography of proteins bound to purified metaphase chromosomes assembled in control or Δ CPC *Xenopus* egg extracts containing [³⁵S]-labelled Rcc1 (loading control) and xHP1 α , β or γ (arrowheads). **c**, Autoradiography of proteins bound to purified metaphase

chromosomes assembled in control or Δ CPC extracts containing [³⁵S]-labelled Rcc1 (loading control) and xHP1 α (WT) or chromodomain mutant xHP1 α W57A (mut). Arrowhead indicates [³⁵S]-labelled xHP1 α . **d**, Binding of xHP1 α –GFP to metaphase chromosomes assembled in Δ CPC extract is competed by an H3K9me3 peptide, but not an unmodified control peptide. DNA was stained with Hoechst 33258 and HP1 α –GFP was visualized by immunostaining using an anti-GFP antibody. Scale bar, 10 μ m.

amount of xHP1 α W57A on metaphase chromosomes purified from Δ CPC extracts was reduced to approximately 20–30% of that seen with the wild-type protein (Fig. 6c). Second, we investigated whether binding of xHP1 to metaphase chromosomes in Δ CPC extracts could be inhibited by a peptide that competes with the chromodomain–H3K9me3 interaction. Although neither an unmodified nor a K9me3 H3-tail peptide had an effect on the chromosomal binding of xHP1 α –GFP in control extracts (Fig. 6d, left), the addition of an H3K9me3 peptide, but not an unmodified control peptide, to Δ CPC extracts significantly reduced the chromosomal binding of xHP1 α –GFP ($\sim$ 50%, $P < 0.0001$, $n = 50$ in each of two independent experiments) (Fig. 6d, right). These results imply that HP1 proteins can only bind to metaphase chromosomes in the absence of H3S10 phosphorylation, and that this association depends on chromodomain–H3K9me3 interaction.

Discussion

Although histone H3S10ph is widely seen as a hallmark of mitosis, the function of this modification during M phase has been enigmatic²⁸. Our data suggest that phosphorylation of H3S10 by Aurora B disrupts the chromodomain–H3K9me3 interaction (Figs 2 and 3), which is important for HP1 recruitment to chromatin during interphase^{11,13}. This disruption causes a net shift in the dynamic HP1–chromatin binding equilibrium towards the unbound state^{7,9,38}. In this reaction sequence, dephosphorylation of H3S10 at the end of mitosis²⁸ re-establishes the overall association of HP1 with chromatin. We propose that this binary ‘methyl/phos switching’ permits dynamic control of the HP1–H3K9me interaction³⁹.

Intriguingly, the mechanism for HP1 release from M-phase chromatin does not involve a temporary loss of H3K9me3 (Fig. 1), but instead requires a combination of this unchanging mark and the dynamic H3S10ph modification that is only transiently added to chromatin during mitosis. We reason that stable transmission of the heterochromatin-defining H3K9me3 mark is needed to accurately convey, from one cell generation to the next, which regions of the genome are supposed to be permanently silenced. If removal of HP1 from M-phase chromatin were accomplished by H3K9me3-erasing demethylase activities, the epigenetic information underlying this mark- and effector-system would have to be accurately re-established at the end of every cell cycle.

In addition to H3S10 phosphorylation, other mechanisms might be involved in the mitotic release of HP1 from chromatin. These might include further modifications of the H3-tail (for example by acetylation on K14, ref. 25), HP1 proteins and/or their interaction partners⁶. Nevertheless, inhibition (Fig. 5), knockdown (Supplementary Fig. S8) or depletion (Fig. 6) of Aurora B is sufficient to cause aberrant interaction of all HP1 isoforms with mitotic, condensed chromatin. Although we cannot exclude the possibility that HP1 proteins themselves might be *in vivo* targets of Aurora B kinase activity (for example, we reproducibly observed increased association of the xHP1 α W57A mutant protein with metaphase chromosomes assembled in Δ CPC extracts, see Fig. 6c), it is known that the phosphorylation level of HP1 β and γ does not increase during mitosis⁷. As phosphorylation of an H3K9me3 peptide is sufficient to dissociate HP1 from this site *in vitro* (Figs 2, 3), we conclude that Aurora B-mediated phosphorylation of H3S10 must be the central event in mitotic release of HP1 from chromatin.

Notably, a fraction of HP1 α , but not HP1 β or γ , remains associated with the (peri-)centromeric chromosome region⁹ (Fig. 5), where it performs important functions for centromere cohesion and kinetochore formation^{10,40} and might be required to identify and define this specialized area of heterochromatin throughout the cell cycle. This mitotic retention of HP1 α at centromeres depends on a carboxy-terminal region of the protein, but is independent of the chromodomain⁸. We therefore suggest that ‘methyl/phos switching’ uniformly disrupts HP1–chromatin interaction but that mechanisms other than chromodomain–H3K9me3

interaction are responsible for the lingering HP1 α association with pericentromeric regions.

What is the function of HP1 dissociation from chromatin during M phase? It is tempting to speculate that removal of HP1 is important for allowing access by factors necessary for mediating proper chromatin condensation and faithful chromosome segregation during mitosis. Indeed, inhibition of Aurora B in vertebrate cells results in defects in chromosome alignment, segregation, chromatin-induced spindle assembly and cytokinesis^{29,35,36,41,42}. Furthermore, mutation of H3S10 causes faulty chromosome segregation in *Tetrahymena* and *S. pombe*, organisms that rely on HP1 and H3K9me3 for the establishment and maintenance of heterochromatin^{43,44}, but not in *Saccharomyces cerevisiae*, an organism that lacks this silencing system⁴⁵. Interestingly, most histone phosphorylation sites are rapidly phosphorylated early in M phase⁴⁶. It remains to be seen whether these bursts in histone phosphorylation are directly involved in the release of proteins bound to interphase chromatin, which might need to be removed to ensure faithful progression through mitosis. It is conceivable that similar ‘methyl/phos switches’ play critical roles in governing other histone–non-histone or even non-histone–non-histone interactions.

METHODS

In vitro protein binding assays were performed as previously described⁴⁷. Details of recombinant protein expression and purification, peptides, and kinase reaction conditions can be found in Supplementary Methods. Western blotting and immunofluorescence analyses were done according to standard protocols (see Supplementary Table S1 for antibodies and dilutions). Anti-xINCENP antibodies³⁶ were used to deplete all components of the CPC from meiotic metaphase II-arrested *Xenopus laevis* egg extracts³⁴. Replicated metaphase chromosomes were assembled and purified as previously described³⁶. Details of association and competition experiments can be found in Supplementary Methods.

Received 3 August; accepted 16 September 2005.

Published online 12 October 2005.

1. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
2. Khorasanizadeh, S. The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259–272 (2004).
3. Grewal, S. I. & Moazed, D. Heterochromatin and epigenetic control of gene expression. *Science* **301**, 798–802 (2003).
4. Grewal, S. I. & Elgin, S. C. Heterochromatin: new possibilities for the inheritance of structure. *Curr. Opin. Genet. Dev.* **12**, 178–187 (2002).
5. Li, Y., Kirschmann, D. A. & Wallrath, L. L. Does heterochromatin protein 1 always follow code? *Proc. Natl Acad. Sci. USA* **99** (suppl.), 16462–16469 (2002).
6. Eissenberg, J. C. & Elgin, S. C. The HP1 protein family: getting a grip on chromatin. *Curr. Opin. Genet. Dev.* **10**, 204–210 (2000).
7. Minc, E., Allory, Y., Worman, H. J., Courvalin, J. C. & Buendia, B. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. *Chromosoma* **108**, 220–234 (1999).
8. Hayakawa, T., Haraguchi, T., Masumoto, H. & Hiraoka, Y. Cell cycle behaviour of human HP1 subtypes: distinct molecular domains of HP1 are required for their centromeric localization during interphase and metaphase. *J. Cell Sci.* **116**, 3327–3338 (2003).
9. Schmiedeberg, L., Weissbart, K., Diekmann, S., Meyer Zu Hoerste, G. & Hemmerich, P. High- and low-mobility populations of HP1 in heterochromatin of mammalian cells. *Mol. Biol. Cell* **15**, 2819–2833 (2004).
10. Pidoux, A. L. & Allshire, R. C. Kinetochore and heterochromatin domains of the fission yeast centromere. *Chromosome Res.* **12**, 521–534 (2004).
11. Stewart, M. D., Li, J. & Wong, J. Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Mol. Cell Biol.* **25**, 2525–2538 (2005).
12. Peters, A. H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
13. Thiru, A. *et al.* Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J.* **23**, 489–499 (2004).
14. Sims, R. J. III, Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19**, 629–639 (2003).
15. Schotta, G., Lachner, M., Peters, A. H. & Jenuwein, T. The indexing potential of histone lysine methylation. *Novartis Found. Symp.* **259**, 22–37 (2004).
16. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
17. Jacobs, S. A. *et al.* Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J.* **20**, 5232–5241 (2001).

18. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
19. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
20. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
21. Maison, C. & Almouzni, G. HP1 and the dynamics of heterochromatin maintenance. *Nature Rev. Mol. Cell Biol.* **5**, 296–304 (2004).
22. Cheutin, T. *et al.* Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**, 721–725 (2003).
23. Festenstein, R. *et al.* Modulation of heterochromatin protein 1 dynamics in primary mammalian cells. *Science* **299**, 719–721 (2003).
24. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genet.* **30**, 329–334 (2002).
25. Mateescu, B., England, P., Halgand, F., Yaniv, M. & Muchardt, C. Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. *EMBO Rep.* **5**, 490–496 (2004).
26. Syka, J. E. *et al.* Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* **3**, 621–626 (2004).
27. Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997).
28. Prigent, C. & Dimitrov, S. Phosphorylation of serine 10 in histone H3, what for? *J. Cell Sci.* **116**, 3677–3685 (2003).
29. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
30. Yasui, Y. *et al.* Autophosphorylation of a newly identified site of Aurora-B is indispensable for cytokinesis. *J. Biol. Chem.* **279**, 12997–13003 (2004).
31. Honda, R., Korner, R. & Nigg, E. A. Exploring the functional interactions between Aurora B, INCENP, and survivin in mitosis. *Mol. Biol. Cell* **14**, 3325–3341 (2003).
32. Chen, J. *et al.* Survivin enhances Aurora-B kinase activity and localizes Aurora-B in human cells. *J. Biol. Chem.* **278**, 486–490 (2003).
33. Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991).
34. Andrews, P. D., Knatko, E., Moore, W. J. & Swedlow, J. R. Mitotic mechanics: the auroras come into view. *Curr. Opin. Cell Biol.* **15**, 672–683 (2003).
35. Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161**, 281–294 (2003).
36. Sampath, S. C. *et al.* The chromosomal passenger complex is required for chromatin-induced microtubule stabilization and spindle assembly. *Cell* **118**, 187–202 (2004).
37. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
38. Sugimoto, K., Tasaka, H. & Dotsu, M. Molecular behaviour in living mitotic cells of human centromere heterochromatin protein HPL α ectopically expressed as a fusion to red fluorescent protein. *Cell Struct. Funct.* **26**, 705–718 (2001).
39. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
40. Obuse, C. *et al.* A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nature Cell Biol.* **6**, 1135–1141 (2004).
41. Ditchfield, C. *et al.* Aurora B couples chromosome alignment with anaphase by targeting BubR1, Mad2, and Cenp-E to kinetochores. *J. Cell Biol.* **161**, 267–280 (2003).
42. Lampson, M. A., Renduchitala, K., Khodjakov, A. & Kapoor, T. M. Correcting improper chromosome–spindle attachments during cell division. *Nature Cell Biol.* **6**, 232–237 (2004).
43. Mellone, B. G. *et al.* Centromere silencing and function in fission yeast is governed by the amino terminus of histone H3. *Curr. Biol.* **13**, 1748–1757 (2003).
44. Wei, Y., Yu, L., Bowen, J., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**, 99–109 (1999).
45. Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291 (2000).
46. Dai, J., Sultan, S., Taylor, S. S. & Higgins, J. M. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes Dev.* **19**, 472–488 (2005).
47. Jacobs, S. A., Fischle, W. & Khorasanizadeh, S. Assays for the determination of structure and dynamics of the interaction of the chromodomain with histone peptides. *Methods Enzymol.* **376**, 131–148 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to M. A. Jelinek and colleagues at Upstate Biotechnologies for developing the monoclonal dual-mark combination-specific anti-H3K9me3S10ph antibody, to S. Hake and C. Barber for purifying H3 for mass spectrometry analysis, and to S. Mollah for initial mass spectrometry analyses. We thank T. Kapoor and Boehringer Ingelheim for providing hesperadin, S. Taylor for the anti-Aurora B antibody, and P. Hemmerich for the HP1-GFP expression constructs. We are grateful to S. Khorasanizadeh and S. Jacobs for their input and help with interpretation of structural data, and to S. Sampath and E. Zeleneova for their input at early stages of this work. This work was funded by grants from the National Institutes of Health (C.D.A. and D.H.F.) and by The Rockefeller University (C.D.A. and H.F.). H.F. is supported by a Searle Scholarship, the Alexandrine and Alexander Sinsheimer Fund, and the Irma T. Hirsch/Monique Weill-Caulier Trust. W.F. is a Robert Black fellow of the Damon Runyon Cancer Research Foundation. H.L.D. is supported by a predoctoral fellowship from the Boehringer Ingelheim Foundation and B.S.T. is supported by an NRSA Training Grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.F. (fischlw@rockefeller.edu) or H.F. (funabih@rockefeller.edu).

Impact origin of sediments at the Opportunity landing site on Mars

L. Paul Knauth¹, Donald M. Burt¹ & Kenneth H. Wohletz²

Mars Exploration Rover Opportunity discovered sediments with layered structures thought to be unique to aqueous deposition and with minerals attributed to evaporation of an acidic salty sea. Remarkable iron-rich spherules were ascribed to later groundwater alteration, and the inferred abundance of water reinforced optimism that Mars was once habitable. The layered structures, however, are not unique to water deposition, and the scenario encounters difficulties in accounting for highly soluble salts admixed with less soluble salts, the lack of clay minerals from acid-rock reactions, high sphericity and near-uniform sizes of the spherules and the absence of a basin boundary. Here we present a simple alternative explanation involving deposition from a ground-hugging turbulent flow of rock fragments, salts, sulphides, brines and ice produced by meteorite impact. Subsequent weathering by intergranular water films can account for all of the features observed without invoking shallow seas, lakes or near-surface aquifers. Layered sequences observed elsewhere on heavily cratered Mars and attributed to wind, water or volcanism may well have formed similarly. If so, the search for past life on Mars should be reassessed accordingly.

Sediment layers discovered by Mars Exploration Rover (MER) Opportunity have been interpreted as siliclastic material deposited by highly acidic waters, which then evaporated to produce Ca/Mg sulphates, chlorides, bromides and jarosite¹. Haematite-rich spherules are then thought to have subsequently formed in the subsurface as concretions, and large crystals were dissolved to produce voids². The strata are considered to represent a once habitable environment by analogy with terrestrial extremophiles inhabiting the acid waters of the Rio Tinto River, Spain². Evidence for deposition in an aqueous system was declared “conclusive”³ and the result has been widely hailed as a milestone in humankind’s search for life elsewhere in the Universe⁴. However, deposition from the surge ejecta of a large meteorite impact is a simple alternative interpretation that accounts for all the observed features and avoids previously unrecognized problems with the aqueous deposition scenario.

Impact surge

Surges are density currents composed of crater ejecta and gases that sweep radially over the substrate away from an explosive crater during its formation (Fig. 1). They are well-known for near-surface tests of chemical and nuclear explosives^{5,6}, described in detail for volcanic explosions⁷, predicted for large planetary impact structures⁸, and found in ejecta from the Chicxulub impact structure⁹. They form layered and cross-bedded deposits that extend radially up to several crater radii away from a crater rim. Emplaced by multiphase, granular flow and influenced by shock wave propagation, the surge and its mechanism vary depending upon the type of cratering event. For impact craters, surges hypothetically originate by winnowing of fine ejecta from the ballistic curtain¹⁰, secondary debris mobilized by ballistic ejecta¹¹, by secondary vapour explosions caused by interaction of residual impact melt and saturated target rocks¹², and by density currents formed from impact breccia and late-stage or distal ejecta¹³. Surge deposition should be an important feature of impact craters on Mars because of atmospheric influence on ejecta

dispersion and the abundant water, ice and/or brine in the regolith^{12,14,15}. Supporting this are the distinctive, ‘fluidized’ ejecta deposits surrounding many martian impact craters (‘rampart’ craters) that display textural features indicative of surge-like flow⁷. Important aspects of surge transport include its ability to deposit ejecta over a larger area than that typical of continuous ballistic ejecta^{16,17}, its deposition of multiple ejecta layers that resemble aeolian or water-laid strata⁷, the wide range of dune-like structures deposited⁷, the extensive production of spherules by accretion of dust-sized particles¹⁸, and the effects of condensation of components of the gas phase^{7,19}.

The high-albedo, cross-laminated stratum at the Opportunity landing site is visible in orbital photographs, and apparently extends over tens of thousands of square kilometres (ref. 2). A key issue is the size of an impact crater that could produce such a large surge deposit. Garvin *et al.*¹⁷ used MOLA (Mars Orbiter Laser Altimeter) data to



Figure 1 | Nevada Test Site nuclear test explosion that produced crater Sedan. The cloud of suspended particles expanding outward along the ground surface is the surge. This surge left a cross-bedded sand deposit up to 1 m thick⁵.

¹Department of Geological Sciences, Arizona State University, Box 871404, Tempe, Arizona 85287-1404, USA. ²Earth and Environmental Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

investigate ejecta thicknesses around martian impact craters with respect to the ejecta thickness function (ETF), $t_e = a(r/R_a)^b$, where t_e is the ejecta blanket thickness, r is the radial distance, R_a is the apparent crater radius, and a and b are fit constants. The ETF is based on scaling studies of explosive and impact craters¹⁶, and thus includes a basic scaling of gravitational and atmospheric effects. Their results show a high variability of the exponent b , with $-24 < b < -0.3$ for craters bisected by MOLA data, which contrasts to the $b = -3 \pm 0.5$ commonly cited for impact craters⁸. For craters surrounded by lobate ejecta blankets of relatively flat profile, they found b to be less than -2 , with cited examples of $b = -1$ and $b = -0.7$. Accordingly, they interpret ejecta profile flatness as a measure of ejecta fluidization and target volatile content. Data for terrestrial volcanic surge deposits^{7,20} yield a best-fit ETF value of $b = -1$ in support of the Garvin *et al.* interpretation (Fig. 2a). Extrapolating this volcanic surge ETF to ejecta deposits for hypothetical 100- and 400-km-diameter impact craters on Mars (Fig. 2b) shows >1 m thicknesses extending up to 600 km from the crater rim.

Rigorous relationships developed specifically for surge deposits have not been published, although the effects of gravity and atmosphere have been applied to martian ballistic ejecta¹¹. Surge transport is governed by transformation of gravitational potential energy to translational kinetic energy⁶ and by creation and dissipation of turbulence⁷. Because gravitational acceleration drives surge run-out but also causes deceleration during frictional contact with the substrate, the scaling of gravity is not expected to be critical for considerations of surge dispersal. However, turbulence is created

from drag that is directly proportional to atmospheric density and has greatest influence at transonic speeds for ejecta moving through the atmosphere. The interplay of a lower rate of turbulence creation for subsonic speeds but higher Mach numbers for martian ejecta remains unexplored.

Mitchell *et al.*²¹ characterized 85 impact crater forms, and found that ejecta deposit diameter is larger than crater diameter by a factor of 2.7 to 4.5, representing a radial extent of about 1 to 2 crater diameters from the crater rim. Overall, these results suggest that for large impact craters (1) surge ejecta deposits show little change in thickness ($<10\%$) over distances of 100 km; (2) surge deposits greater than 2 m in thickness might extend over 500,000 km² from a single impact crater; and (3) observed impact surge deposit extents are likely to be less than their initial range because of erosion or late-stage non-newtonian flow behaviour displayed as pedestal or rampart formation.

It is thus possible that the layers traversed by the rover resulted from one impact event, possibly the 450-km-wide crater Schiaparelli lying about 2 crater diameters to the east. Alternatively, they could be made up of multiple, interlaced surge deposits from numerous smaller and closer craters. In this case, the light-coloured layer identified from orbit is not a single chronostratigraphic unit, but rather owes its light colour to sulphate wicking as described below. In any case, impact surge is a reasonable mechanism for thinly layered deposits on Mars, even at great distances from the source craters.

Bedding and cross-bedding

Bedding and sedimentary structures created by surge closely resemble those produced by aeolian and subaqueous deposition. Because of the complexity of multiphase granular flow, a remarkably wide range of depositional conditions develop that are dimensionally analogous to other sedimentary environments. For example, surges may be erosive or depositional, such that channelling and delta-like fore-set bedding results. Multiphase sound speed can be as low as several tens of metres per second, such that standing shock waves exist and create effects similar to those of the transition from supercritical to subcritical flow in aqueous conditions. Cross-bedding and other sedimentary structures can also be created from impact ejecta that clump together in the atmosphere and flow to the ground as density currents²².

Festoon cross-bedding observed in Eagle crater was interpreted as uniquely subaqueous in origin⁷. However, such cross-bedding occurs in terrestrial surge deposits^{23,24} (Fig. 3). The long, low-angle cross-beds in the darker layers underlying the high-albedo rim unit at Endurance crater (Fig. 4a) are common in surge deposits (Fig. 4b), as are the high-angle cross-beds that underlie these. The combined thickness of the observed stratified units is of the order of several metres, perhaps consistent with multiple surge events. A single surge event can emplace a deposit of multiple layers²³, and a single impact event might produce multiple surges if vapour explosions continue after formation of the transient crater¹².

Origin of salts

A serious flaw in the evaporating lake scenario is that the most soluble salts (halides, Mg-sulphate) occur together with the least soluble salts (Ca-sulphate, jarosite). This does not happen in evaporating water masses where the least soluble salts precipitate first and line the coastal areas. The most soluble salts form in residual brine pools in the lowest areas, usually towards the centre of the basin²⁵. Bromide, in particular, is so soluble that bromide-rich evaporites should never precipitate together with sulphates.

Squyres *et al.*¹ suggest that basaltic material was weathered in surrounding areas, transported to the site by water, and deposited together with sulphates and jarosite during evaporation of the inflowing acid waters. However, acid rivers, lakes, and aquifers on and within a basaltic substrate cannot be sustained because basalt reacts rapidly with acid, particularly basalt in the form of glass and

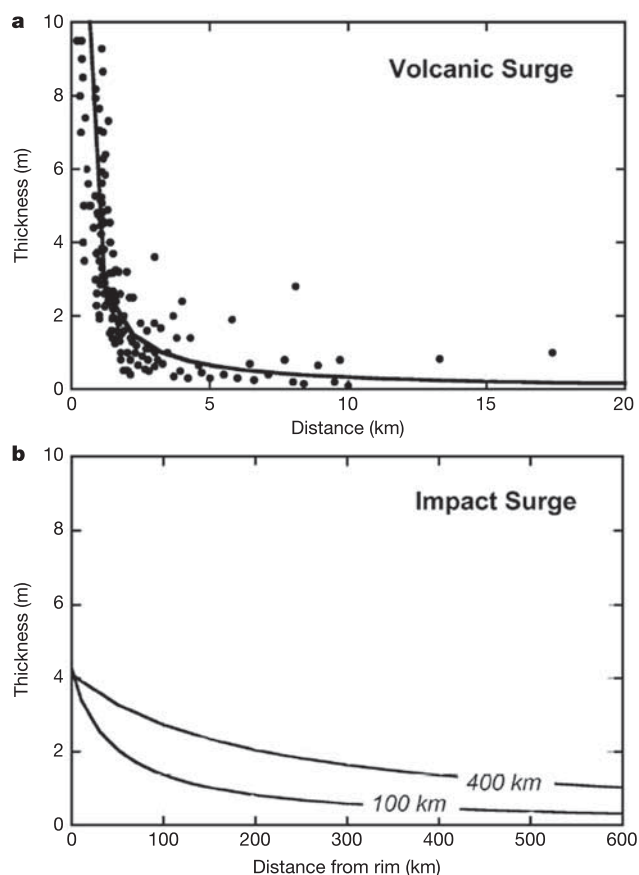


Figure 2 | Plots of surge deposit thickness versus radial distance. **a**, The variation in thickness for terrestrial volcanic surge deposits is plotted as a function of distance from crater centre (points beyond 20 km are not shown). **b**, The best-fit Ejecta Transfer Function from the volcanic data are used to estimate the thickness of impact surge deposits for 100- and 400-km-diameter craters on Mars.

fragmental debris in the martian regolith. Clay minerals are produced and the solution is neutralized²⁶. The apparent presence of only trace amounts of clay minerals on Mars³ thus contradicts scenarios involving large acid water masses and, particularly, acid ground waters. The terrestrial Rio Tinto River analogy is not ideal, because the sediment from this river has voluminous clay minerals and the acid level is enhanced by over 3,000 years of human mining activity on the Earth's largest-known volcanogenic sulphide deposit. No such clay deposits, upland massive sulphide source, upland drainage channels, or deltas extending into the putative lake have been observed near Meridiani Planum. The lack of any visible shorelines or other basin boundary (ref. 27, and references cited therein) is an additional problem.

An alternative explanation for the widespread presence of sulphates and chlorides mixed with basaltic fragments arises as a

consequence of the previously known evidence for a relatively brief period early in martian history when large amounts of water flowed over the surface (ref. 28, and references cited therein). Up to 90% of this water was lost from the planet, based on the current D/H ratio of the atmosphere (ref. 29, and references cited therein). Loss of this water from the atmosphere would necessarily have caused the remaining hydrosphere to become evapoconcentrated into a brine¹⁵. This brine, lodged in the megaregolith, would necessarily have reacted with the basaltic materials and evolved into Ca-Mg-Na-Cl⁻-rich brine. With the onset of global freezing, the brine in the megaregolith would have necessarily undergone fractional freezing to produce water ice, chloride, bromide and sulphate salts, and highly concentrated, eutectic brine with freezing points below current martian equatorial temperatures³⁰. Subsequent large impacts into the megaregolith would scatter not only basaltic materials over large distances, but also the included salts, ice and brine. In the surge scenario, a large impact might excavate through the entire megaregolith and thus eject the full complement of phases originally separated via evapoconcentration and fractional crystallization. Phases expected in a martian impact surge would therefore be sand-sized and finer basalt fragments and glasses together with a disequilibrium mechanical mixture of hydrohalite, antarcticite, Mg-Ca sulphates, minor clays, chlorides, bromides, ice, brine, and minor exotic salts formed during fractional crystallization.

The total thickness of deposits at Meridiani Planum could be 1 km,

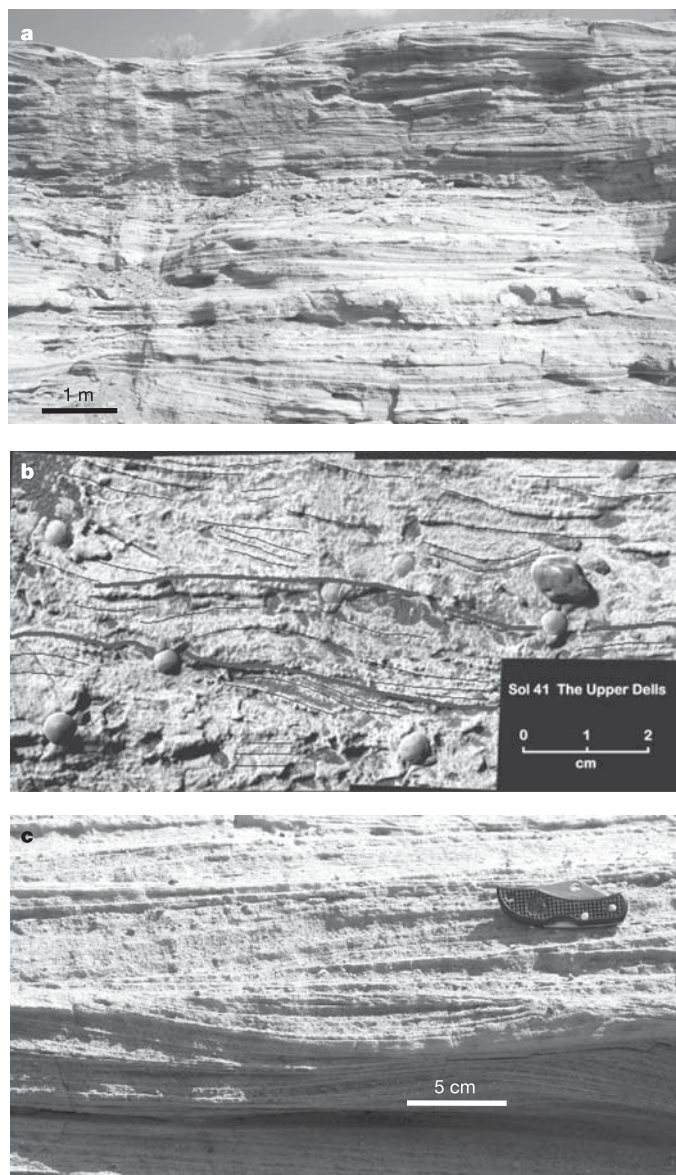


Figure 3 | Terrestrial surge deposits compared with cross-stratified martian deposits. **a**, Typical layered and cross-bedded aspect of a terrestrial deposit, Kilbourne Hole, New Mexico. **b**, Upper Dells mosaic taken on sol (martian day) 41. Lines added by the MER team to highlight cross-sets. **c**, Festoon cross-beds from Kilbourne Hole, New Mexico. Festoon cross-sets in terrestrial surges occur at the same scale as those observed on Mars and need not imply an aqueous origin.

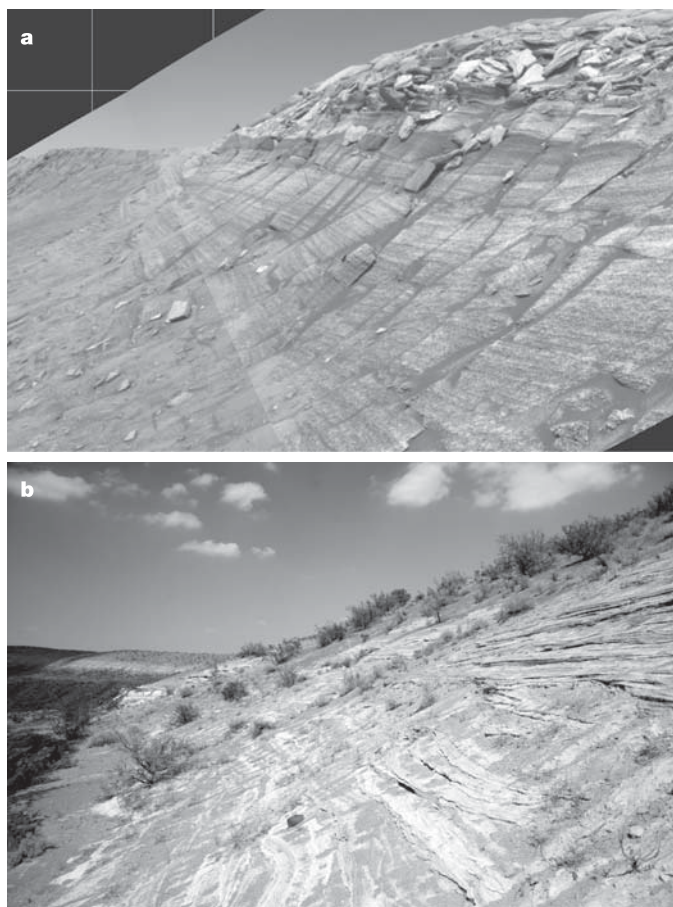


Figure 4 | Martian strata compared with terrestrial surge strata. **a**, Wall of Endurance crater on Mars, showing long, low-angle cross-sets overlying high-angle cross-sets (upper left part of photograph). The sloping straight line is an artefact of image stitching. The bedding displayed here is common in surge deposits. Impact surge explains all stratification in terms of only one process. **b**, Outcrop appearance of typical, layered surge deposits, Kilbourne Hole, New Mexico.

or more^{31,32}. An evaporite origin is unlikely if all are rich in sulphate because the amount of water implied is untenable. For example, evaporation of a 1 km column of terrestrial sea water yields less than 2 m of sulphate. Stacked surge deposits from numerous distant craters could be sulphate-rich, and might occur interbedded with volcanic ash and wind-blown deposits. In any case, only the topmost layers are visible to MER and these strongly resemble surge deposits.

Following emplacement, the heterogeneous jumble of mechanically emplaced phases would necessarily undergo diagenetic reorganization following mobilization of small amounts of interstitial waters and water films. Most of the salts are hygroscopic and/or deliquescent, and would absorb water vapour from interstitial cavities, ice and the overlying atmosphere. Mixed with melted ice from warm periods, local brine pockets would yield early diagenetic crystal growth, including that resulting from Ostwald ripening³³, where smaller crystals scattered through the mechanical mixture yield to larger ones. Such mineralogical stabilization after deposition is inevitable, especially considering that >3 billion years were available.

Crystal moulds visible in the upper light-coloured layer have been interpreted as monoclinic crystals, possibly gypsum, that formed diagenetically and were subsequently dissolved². The rock matrix is inferred to have at present up to 30–40% finer-grained sulphate². It is highly unlikely that the largest crystals would dissolve before smaller crystals of the same mineral. It is more likely that these larger crystals were a soluble halide such as hydrohalite, a monoclinic mineral that would probably have grown in briny pore fluids from the mechanically emplaced salt fragments via Ostwald ripening. With time and addition of minor water, brines would migrate downward and carry the most soluble components deeper into the sedimentary pile. Chloride and bromide salts would be preferentially removed and the less soluble sulphates left behind. Deliquescent phases would be removed by ice melt during warm periods and/or would absorb water vapour, become fluid, and drain out. This simpler scenario for the moulds does not require larger crystals to dissolve preferentially to smaller ones, and does not require introduction of late, fresh ground

waters. In the exceedingly dry atmosphere of Mars, further near-surface concentration of sulphates could occur by 'wicking up' processes similar to those that produce sulphate efflorescences on desert mine dumps on Earth.

Differential solubility thus explains the observed cavities and the relatively abundant sulphate as a lag and/or efflorescence following downward movement of the more soluble Cl. The entire mass need not have been bathed in water. Instead, there were films, pockets and preferred flow paths. Small areas that stayed relatively dry retained more of the initial composition that was rich in Cl and Br. This scenario also explains the paucity of observed clay minerals because only small amounts of neutral water were involved. The absence of clay minerals and playa mud layers is a serious problem for the evaporating acid lake hypothesis but not for the impact surge hypothesis.

Spheroids. Abundant well-sorted, largely spherical, haematite-rich grains about 5 mm in diameter occur as several per cent of the bedrock and accumulated as a lag on the surface². Occasional banding and grooves parallel to bedding, termination of crystal moulds against the spherules, their uniform distribution, the general lack of bedding disturbance, and the presence of rare, apparent doublets have been advanced as evidence that the spherules are concretions that formed diagenetically in the phreatic zone of a groundwater system². Concretions in the Navajo Sandstone have been invoked as a possible terrestrial analogue³⁴.

Although terrestrial concretions with almost perfect sphericity can occur, the size uniformity and high sphericity of the martian spherules are rare or even unknown over outcrop scales on Earth. The Navajo Sandstone concretions come in a wide variety of sizes and shapes, are not distributed uniformly in outcrop, and occur in clusters, zones and irregular, nonlinear clumps of multiple concretions of random size and orientation³⁴. They represent almost pure quartz sand cemented locally by small amounts of goethite, haematite and carbonate. The composition and internal nature of the martian examples are not yet fully known, but they do not resemble terrestrial concretions other than having an iron oxide component

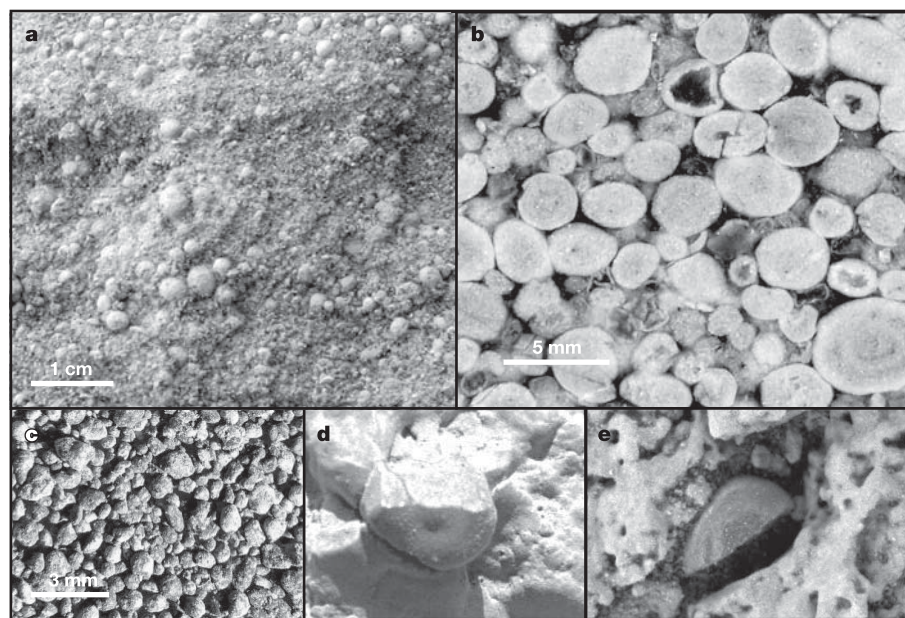


Figure 5 | Martian spherules compared with terrestrial accretionary lapilli and impact spherules. **a**, Terrestrial accretionary lapilli in surge deposit, Kilbourne Hole, New Mexico. These are similar in size, shape and sphericity to martian spherules. **b**, Compacted, terrestrial accretionary lapilli from the 3.5-Gyr-old Onverwacht Formation, South Africa. Initially interpreted as volcanic accretionary lapilli⁵⁰, these may be impact surge accretionary lapilli

because impact spherule beds occur in the same formation⁴⁰ and the two spherule types are locally intermixed. **c**, Iron condensation spherules, Meteor crater, Arizona. **d**, Broken, cored 4-mm spherule on Mars, sol 28. **e**, Broken spherule on Mars, sol 142. Spherule diameter is approximately 4 mm and appears to have layered concentric structure similar to terrestrial accretionary lapilli.

and being spheroidal. The martian spherules carry the orbital Thermal Emission Spectrometer (TES) haematite signature and are thus inferred to cover an area of 150,000 km (refs 2, 31). If concretions, they grew uniformly in a vast shallow unfrozen aquifer on a scale probably unknown in the rock record on Earth, which has always had widespread aquifers and abundant concretions of diverse mineralogies and forms.

On Earth, volcanic surge deposits commonly contain accretionary lapilli (Fig. 5a), as well as other small spherical particles, all of which have been found in terrestrial impact ejecta^{35,36}. Accretionary lapilli^{18,37} form largely owing to moisture within the moving surge cloud that causes grains to adhere one to another, building up concentric layers while in turbulent motion, much in the way of hailstones. The lapilli preferentially gather fine dust but do not begin formation until steam condensation has begun. Where iron-rich particles are accreted in these fumarole-like conditions, iron is rapidly oxidized³⁸. Once lapilli have grown large enough, they fall out of the surge cloud. At any one location within the surge deposit, their size is therefore fairly uniform. Spherical ejecta are also hydrodynamically formed in certain fragmentation modes of water/melt interaction³⁹.

Large impacts are known to produce condensation spherules (including doublets) of similar size to the lapilli⁴⁰. Regional sheets of accretionary lapilli and impact condensation spherules occur in terrestrial Archaean strata (Fig. 5b), often mixed in the same horizon and difficult to distinguish. The large geographic areas over which the terrestrial spheroids occur are comparable to the area yielding the orbital TES haematite signature. Well-sorted spheroids may thus have been produced from impact events and then deposited by a surge or more distal density currents²²; a regional aquifer is not necessary.

Meridiani Planum was selected as a landing site primarily because it is one of the few places displaying an extensive haematite infrared emission signature. In terms of our hypothesis, the TES signature itself may represent oxidized iron in accretionary lapilli made up largely of basaltic glass particles. However, the uniqueness of the deposit may indicate that the impactor was a large iron meteorite that yielded a large population of iron condensation spherules as well. Iron meteorites are rare and large iron meteorites that could produce a >100 km crater are rarer still. Iron condensation spherules are produced during iron meteorite impact events (Fig. 5c). Oxidation of the iron to haematite during volatilization/condensation is most likely, but later weathering could also account for the enhanced haematite signal visible from orbit. Chemical analyses indicate a good correlation between Ni and Fe when comparing spherule-rich and non-spherule-rich areas⁴¹. This is problematical for concretions, but is to be expected if there was a major nickel-iron meteorite component in the spherules. This site was specifically chosen because of its peculiar TES signature, so landing on a spot unusually rich in condensation spherules from an iron impactor would not be so fortuitous.

Spherules are typically girdled with rinds of matrix that eventually abrade away upon release from the rock. Remnant ridges and grooves are thus not compelling evidence for a replacement origin; cross-sections of the spherules made by the Rotary Abrasion Tool are typically circular. The large crystals that abut the spheroids are compatible with our scenario of impact surge deposition followed by diagenetic crystal growth. This scenario does not require removal of siliciclastic grains to make room for the spherules, which is something required in the concretion scenario. Accretionary lapilli and condensation spherules do not necessarily disrupt sedimentary lamination in known terrestrial examples, so the usual lack of such disruption on Mars need not argue for concretion growth. Some crystals terminate against spheroids, but this can happen during growth and does not imply truncation² indicative of later spheroid growth. At least two fractured and wind-polished concretions display possible concentric zonation similar to that in many terrestrial

accretionary lapilli (Fig. 5d, e). In any case, impact spherules and accretionary lapilli are commonly internally uniform and are compatible with all imagery returned so far.

Jarosite and other features

About 28% of the Fe in the rock outcrops is inferred to be in jarosite⁴². This mineral develops in arid regions on Earth when rocks bearing iron sulphide minerals are mined, crushed and exposed to weathering⁴³. The fresh sulphides react rapidly with oxygen and small amounts of water to produce films and rivulets of acid sulphate, which immediately react with the silicate host rocks to produce clays and coatings of jarosite. Following rains, jarosite is eventually converted into goethite or haematite via incongruent dissolution. Mixtures of jarosite, goethite and haematite thus occur together in mine dumps and tailings that contain sulphide. An impact surge deposit on Mars might have three sources of sulphur sufficient to account for jarosite alteration: (1) Fe-reduction of evaporitic sulphate to sulphide during impact devolatilization, (2) primary sulphide concentrations near the bottoms of impact excavated mafic magma chambers and flows⁴⁴, and (3) sulphide (1–2 wt% S) contained in the iron bolide itself. Subsequent martian oxidative and frost weathering over billions of years could be comparable to terrestrial arid-region weathering. Minor water produces films of sulphuric acid, which are then partly neutralized by reaction with the pulverized rocks to produce localized coatings of jarosite. The co-existing haematite represents material that was not contacted by the acid films or where continued addition of small amounts of water leached the acid component of jarosite. Such surface and near-surface coatings of jarosite could cause this mineral to appear more plentiful than it actually is.

The Meridiani deposit is remarkable for its regional flatness and its paucity of ballistic ejecta blocks. The extreme flatness is characteristic of lake deposits, but it is also characteristic of the tops of terrestrial volcanic surge deposits. These flow with negligible yield strength and thicken over topographically low areas and pond within them, leading to a tendency to form flat surfaces⁴⁵. Topographic flatness is therefore compatible with either a lake deposit or an impact surge deposit. The lack of later ballistic ejecta blocks is problematic in all scenarios if the age of the deposit is several billion years as inferred¹. A younger age is no problem for the impact surge hypothesis but is difficult for the lake hypothesis because it would require standing bodies of water later than normally inferred. It is also possible that this area simply escaped heavy mantling by ballistic impact ejecta.

Possible desiccation cracks have been observed in Endurance crater and were attributed to wetting/drying episodes. Desiccation of inherently wet volcanic surge deposits on Earth is common⁴⁶. The martian examples could thus represent drying out of wet impact surge deposits.

Discussion

The impact surge hypothesis avoids the contradictions implied by aqueous deposition and accounts for all the features observed by the MER Opportunity with a minimum number of events and processes. The scenario calls for a large iron meteorite impact (crater probably >100 km) into a megaregolith containing salts, ice and brine. The enormous wet surge created by this impact, together with surges from secondary impacts, deposited an extremely flat, distal, cross-stratified and layered mechanical mixture of fine basaltic particles, salts, ice, brine, accretionary lapilli and condensation spherules. Over the next 3 billion years, diagenesis and downward flow of local, thin films and droplets of brine formed by salt deliquescence and dissolution of salts in ice-melt moved the more soluble phases (chlorides) downward while the least soluble salts (sulphates) remained or even wicked towards the surface. Acid sulphates formed by oxidation of sulphides in the excavated fragments could have produced jarosite coatings. In this scenario, the discoveries at Meridiani Planum do provide additional evidence of an early

evapoconcentrated hydrosphere on Mars, but this hydrosphere had already disappeared into the megaregolith when a large impact produced this surge deposit. This scenario is wholly compatible with the post-Noachian cold and dry environment currently advocated for Mars based on data from the landing site of Spirit, the other MER⁴⁷.

Although our discussion focuses on the Opportunity site, we note that meteorite impact provides a possible explanation for many finely layered deposits elsewhere on Mars that have previously been considered in terms only of aqueous, aeolian or volcanic origin. Many of the layered blocks observed at the Spirit site in Gusev crater are similar to what is observed at the Opportunity site with respect to fabrics, textures, mineralogy and chemistry, and may also be impact or volcanic surge deposits. Impacts into megaregolith rich in salts, ice and residual brine can readily account for the widespread distribution of salts in martian surface materials, as observed at previous landing sites and now from orbit⁴⁸. Impact surge deposition carries no particular advantage for microfossil preservation, so the search for past life on Mars should probably be re-directed. Better prospects lie, perhaps, in the tiny fractures in surface rocks everywhere. Small mineral deposits from aqueous films are likely there, some with potential isotopic biosignatures⁴⁹. In any case, impact surge should be seriously considered as an alternative, simple explanation for the origin of this and other layered deposits on Mars.

Received 20 June; accepted 26 October 2005.

- Squyres, S. W. *et al.* The Opportunity Rover's Athena science investigation at Meridiani Planum, Mars. *Science* **306**, 1698–1703 (2004).
- Squyres, S. W. *et al.* In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* **306**, 1709–1714 (2004).
- Christensen, P. R. *et al.* Mineralogy at Meridiani Planum from the Mini-TES experiment on the Opportunity rover. *Science* **306**, 1733–1739 (2004).
- Kerr, R. A. On Mars, a second chance for life. *Science* **306**, 2010–2012 (2004).
- Carlson, R. H. & Roberts, W. A. *Ejecta Studies, Project Sedan* (PNE-217P, University of California Lawrence Radiation Laboratory, Livermore, California, 1962).
- Glasstone, S. & Dolan, P. J. *The Effects of Nuclear Weapons* 3rd edn (US Department of Defense and Energy Research and Development Administration, Washington DC, 1977).
- Wohletz, K. H. in *Developments in Volcanology, from Magma to Tephra, Modelling Physical Processes of Explosive Volcanic Eruptions* (eds Freundt, A. & Rosi, M.) 247–312 (Elsevier, Amsterdam, 1998).
- Melosh, H. J. *Impact Cratering* (Oxford University Press, New York, 1989).
- Dressler, B. O., Sharpton, V. L., Schwandt, C. S. & Ames, D. Impactites of the Yaxcopoil-1 drilling site, Chicxulub impact structure: Petrography, geochemistry, and depositional environment. *Meteorit. Planet. Sci.* **39**, 857–878 (2004).
- Oberbeck, V. R. The role of ballistic erosion and sedimentation in lunar stratigraphy. *Rev. Geophys. Space Phys.* **13**, 337–362 (1975).
- Schultz, P. H. & Gault, D. E. Atmospheric effects on Martian ejecta emplacements. *J. Geophys. Res.* **84**, 7669–7687 (1979).
- Carr, M. H. *et al.* Martian impact craters and emplacement of ejecta by surface flow. *J. Geophys. Res.* **82**, 4055–4065 (1977).
- Osinski, G. R., Grieve, R. A. F. & Spray, J. G. The nature of the groundmass of surficial suevite from the Ries impact structure, Germany, and constraints on its origin. *Meteorit. Planet. Sci.* **39**, 1655–1683 (2004).
- Wohletz, K. H. & Sheridan, M. F. Martian rampart crater ejecta: Experiments and analysis of melt-water interaction. *Icarus* **56**, 15–37 (1983).
- Knauth, L. P. & Burt, D. M. Eutectic brines on Mars: Origin and possible relation of young seepage features. *Icarus* **158**, 267–271 (2002).
- McGetchin, T. R., Settle, M. & Head, J. W. Radial thickness variation in impact crater ejecta: Implications for lunar basin deposits. *Earth Planet. Sci. Lett.* **20**, 226–236 (1973).
- Garvin, J. B., Sakimoto, S. E. H., Frawley, J. J. & Schnetzler, C. Global geometric properties of Martian impact craters. *Lunar Planet. Sci.* **33**, abstr. 1255 (2002).
- Sheridan, M. F. & Wohletz, K. H. in *Microbeam Analysis* (ed. Gooley, R.) 35–38 (Univ. San Francisco Press, San Francisco, 1983).
- Kieffer, S. W. & Simonds, C. H. The role of volatiles and lithology in the impact cratering process. *Rev. Geophys. Space Phys.* **18**, 143–181 (1980).
- Valentine, G. A., Buesch, D. & Fisher, R. V. Basal layered deposits of the Peach Springs Tuff, northwestern Arizona, USA. *Bull. Volcanol.* **51**, 395–414 (1983).
- Mitchell, D. E., Sakimoto, S. E. H. & Garvin, J. B. MOLA topography and morphometry of rampart and pedestal craters, Mars. *Lunar Planet. Sci.* **33**, abstr. 1805 (2002).
- Melosh, H. J. Impact ejecta sedimentation processes in the atmosphere and ocean. *Meteorit. Planet. Sci.* **39**, abstr. 67 (2004).
- Fisher, R. V. & Walters, A. C. Base-surge bed forms in maar volcanoes. *Am. J. Sci.* **268**, 157–180 (1970).
- Fisher, R. V. & Schmincke, H.-U. *Pyroclastic Rocks* 249–256 (Springer, New York, 1984).
- Hunt, C. B., Robinson, T. W., Bowles, W. A., Washburn, A. L. & Hunt, C. *Hydrologic Basin, Death Valley, California* (Professional Paper 494-B, US Geological Survey, Washington DC, 1966).
- Bullock, M. A. & Moore, J. M. Aqueous alteration of Mars-analog rocks under an acidic atmosphere. *Geophys. Res. Lett.* **31**, L14701, doi:10.1029/2004GL019980 (2004).
- Golombek, M. P. *et al.* Selection of the Mars Exploration Rover landing sites. *J. Geophys. Res.* **108**, 8072, doi:10.1029/2003JE002074 (2003).
- Baker, V. R. Water and the Martian landscape. *Nature* **412**, 228–236 (2001).
- Jakosky, B. M. & Phillips, R. J. Mars' volatile and climate history. *Nature* **412**, 237–244 (2001).
- Burt, D. M. & Knauth, L. P. Electrically conducting, Ca-rich brines, rather than water, expected in the Martian subsurface. *J. Geophys. Res.* **108**, 8026, doi:10.1029/2002JE001862 (2003).
- Christensen, P. R. & Ruff, S. W. Formation of the hematite-bearing unit in Meridiani Planum: Evidence for deposition in standing water. *J. Geophys. Res.* **109**, E08003, doi:10.1029/2003JE002233 (2004).
- Hynek, B. M., Arvidson, R. E. & Phillips, R. J. Geologic setting and origin of Terra Meridiani hematite deposit on Mars. *J. Geophys. Res.* **107**, 5088, doi:10.1029/2002JE001891 (2002).
- Boistelle, R. & Astier, J. P. Crystallization mechanisms in solution. *J. Cryst. Growth* **90**, 14–30 (1988).
- Chan, M. A., Breiter, B., Parry, W. T., Ormo, J. & Komatsu, G. A possible terrestrial analogue for haematite concretions on Mars. *Nature* **429**, 731–734 (2004).
- Graup, G. Terrestrial chondrules, glass spherules and accretionary lapilli from the suevite, Ries Crater, Germany. *Earth Planet. Sci. Lett.* **55**, 407–418 (1981).
- Bohor, B. F. & Glass, B. P. Origin and diagenesis of K/T impact spherules—From Haiti to Wyoming and beyond. *Meteoritics* **30**, 182–198 (1995).
- Schumacher, R. & Schmincke, H.-U. Models for the origin of accretionary lapilli. *Bull. Volcanol.* **56**, 626–639 (1995).
- Catling, D. C. & Moore, J. M. The nature of coarse-grained crystalline hematite and its implications for the early environment of Mars. *Icarus* **165**, 277–300 (2003).
- Wohletz, K. H. & McQueen, R. G. Volcanic and stratospheric dust-like particles produced by experimental water-melt interactions. *Geology* **12**, 591–594 (1984).
- Lowe, D. R. *et al.* Characteristics, origin, and interpretation of Archean impact-produced spherule beds, 3.47–3.22 Ga, in the Barberton Greenstone Belt, South Africa: Keys to the role of large impacts on the evolution of the early Earth. *Astrobiology* **3**, 7–48 (2003).
- Brueckner, J. *et al.* Hematite on the surface of Meridiani Planum and Gusev Crater. *Lunar Planet. Sci.* **36**, abstr. 1767 (2005).
- Klingelhofer, G. *et al.* Jarosite and hematite at Meridiani Planum from Opportunity's Moessbauer spectrometer. *Science* **306**, 1740–1745 (2004).
- Ditirzac, J. E. & Jambor, J. L. Jarosites and their application in hydrometallurgy. *Rev. Mineral. Geochem.* **40**, 405–452 (2000).
- Burns, R. G. & Fisher, D. S. Evolution of sulfide mineralization on Mars. *J. Geophys. Res.* **95**, 14169–14173 (1990).
- Cas, R. A. F. & Wright, J. V. *Volcanic Successions Modern and Ancient* (Chapman & Hall, London, 1987).
- McPhie, J., Walker, G. P. L. & Christiansen, R. L. Phreatomagmatic and phreatic fall and surge deposits from explosions at Kilauea volcano, Hawaii, 1790 A.D. *Bull. Volcanol.* **52**, 334–354 (1990).
- Golombek, M. P. *et al.* Climate change from the Mars Exploration Rover landing sites: From wet in the Noachian to dry and desiccating since the Hesperian. *Lunar Planet. Sci.* **36**, abstr. 1539 (2005).
- Gendrin, A. *et al.* Sulfates in Martian layered terrains: The OMEGA/Mars Express view. *Science* **307**, 1587–1591 (2005).
- Knauth, L. P., Brilli, M. & Klonowski, S. Isotope geochemistry of caliche on basalt. *Geochim. Cosmochim. Acta* **67**, 185–195 (2003).
- Lowe, D. R. & Knauth, L. P. The oldest marine carbonate ooids reinterpreted as volcanic accretionary lapilli, Onverwacht Group, South Africa. *J. Sedim. Petrol.* **48**, 709–722 (1978).

Acknowledgements L.P.K. is supported by the NASA Exobiology Program. We thank C. Moore for supplying the iron condensation spherules from Meteor crater, and G. R. Osinski for comments on an early version of the manuscript.

Author Contributions All authors contributed equally to the ideas and interpretations. L.P.K. wrote the initial draft and managed revisions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.P.K. (Knauth@asu.edu).

A volcanic environment for bedrock diagenesis at Meridiani Planum on Mars

Thomas M. McCollom^{1,2} & Brian M. Hynek²

Exposed bedrocks at Meridiani Planum on Mars display chemical and mineralogical evidence suggesting interaction with liquid water^{1–6}. On the basis of morphological observations as well as high abundances of haematite and sulphate minerals, the rocks have been interpreted as sediments that were deposited in a shallow body of briny water with subsequent evaporation leaving behind the sulphate minerals^{1,2}. The iron-sulphur mineralization at Meridiani has also been inferred to be analogous to that produced during oxidative weathering of metal sulphide minerals, such as occurs at acid mine drainage sites⁷. Neither of these interpretations, however, is consistent with the chemical composition of the rocks. Here we propose an alternative model for diagenesis of Meridiani bedrock that involves deposition of volcanic ash followed by reaction with condensed sulphur dioxide- and water-bearing vapours emitted from fumaroles. This scenario does not require prolonged interaction with a standing body of surface water and may have occurred at high temperatures. Consequently, the model invokes an environment considerably less favourable for biological activity on Mars than previously proposed interpretations.

The Mars Exploration Rover (MER) Opportunity has returned close-up images of exposed bedrocks from Meridiani Planum, as well as data on their chemical and mineralogical composition^{1–6}. The most distinguishing features of the bedrock include: (1) spherical haematite nodules (the so-called 'blueberries'), (2) 'festoon' bedforms in the rock structure, (3) high concentrations of sulphur, probably in the form of sulphate minerals, (4) anomalously high Br:Cl ratios, and (5) lath-shaped cavities (vugs) suggesting that crystals of a sulphate mineral such as gypsum were formerly present. On the basis of these observations, the MER science team^{1,2} has interpreted the bedrock as siliciclastic sediments deposited in a shallow body of briny water, with subsequent evaporation leaving behind sulphate minerals. They inferred that the rocks are composed of roughly 50 wt% siliciclastic material derived from basaltic rocks, 40 wt% evaporite sulphates and 10 wt% haematite².

However, the composition of the Meridiani bedrocks indicates that the formation model advocated by the MER team is not plausible (Fig. 1). Ratios of cations including Fe, Mg, Ca and Na to (Si + Al) in the rocks are nearly identical to the basaltic martian meteorite Shergotty^{8,9}, and also very similar to unweathered basaltic rocks at Meridiani⁴ and Gusev crater¹⁰. The compositional data strongly suggest that the Meridiani rocks represent typical martian basalt with a sulphur component added (Fig. 1). Indeed, if the SO₃ abundances of Meridiani bedrocks are reduced to the level of other martian rocks, their composition is nearly identical to that of Shergotty (Supplementary Table S1 and Supplementary Fig. S1).

Consequently, any model for the formation of the Meridiani bedrocks must account for enrichment in S but not in any major cations. If the sulphate were attributable to precipitation of salts from an evaporating brine as suggested by the MER team, the rocks would

be enriched in a balancing cation (for example, Ca, Mg or Fe)—but this is not observed. Oxidative weathering of metal sulphide minerals has also been proposed as an analogue to the mineralogy of Meridiani bedrock⁷, but this is similarly implausible because it cannot account for addition of S without a concurrent increase in some cation (Fe is typically the predominant cation in such environments). Furthermore, there is no evidence for the presence of metal sulphide deposits at or near Meridiani.

There are few mechanisms by which the rocks could be enhanced in sulphur without a concurrent increase in cations. The most likely sources would be SO₂ gas or sulphuric acid. It is unlikely that ground water or surface runoff could provide a source of sulphuric acid without cations, as interaction of acidic fluid with basaltic rocks along flowpaths leading to the bedrock would neutralize the fluid and result in high dissolved concentrations of Mg and Ca. Deposition of sulphuric acid from atmospheric sources, formed by reaction of volcanic SO₂ emissions with water vapour, is a possibility, but the extent of alteration and presence of haematite suggest alteration at

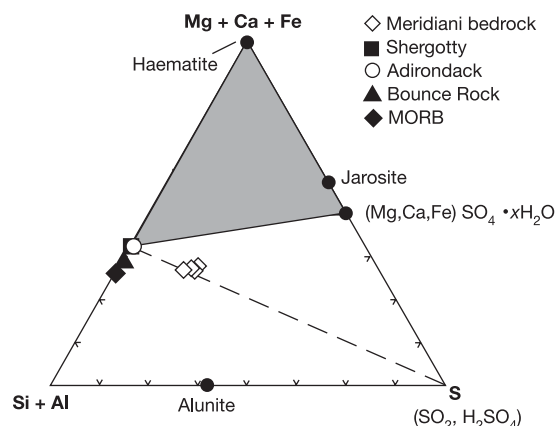


Figure 1 | Ternary diagram comparing the molar chemical compositions of Meridiani bedrocks analysed by Opportunity⁴ with typical martian basalts. Martian basalt compositions are represented by the martian meteorite Shergotty^{8,9}, 'Adirondack'¹⁰ (a basalt analysed by the MER Spirit in Gusev crater), and 'Bounce Rock'⁴ (analysed by Opportunity on Meridiani Planum). Also shown for reference are compositions of typical terrestrial mid-ocean ridge basalt (MORB) and several selected minerals (filled black circles labelled by name). If the bedrocks at Meridiani were composed of a combination of basaltic clasts, haematite, and sulphate salts formed as a sedimentary/evaporite deposit as suggested by the MER team^{1,2}, their composition should fall within the shaded triangle, but this is not observed. Instead, the bedrocks fall on a two-component mixing line between martian basalt and the S endmember (dashed line), providing strong evidence that the rocks' bulk composition represents typical martian basalt to which a pure sulphur component (for example, SO₂, H₂SO₄) has been added.

¹Center for Astrobiology and ²Laboratory for Atmospheric and Space Physics, Campus Box 392, University of Colorado, Boulder, Colorado 80309, USA.

high temperature¹¹. Furthermore, the distribution of sulphate-rich rocks does not appear to be consistent with an atmospheric source. Geologic mapping based on remote sensing indicates that the layered deposits observed at the rover landing site are spatially extensive ($>3 \times 10^5 \text{ km}^2$), and the entire unit appears to be diagenetically related¹². Recent results from the OMEGA visible/infrared mineralogical mapping spectrometer aboard the Mars Express spacecraft indicate that bright outcrops beyond Meridiani Planum (but part of the same geologic unit) are rich in sulphates¹³, strengthening this connection. A topographic trough immediately north of the haematite-bearing plains provides a cross-section of the unit, and reveals complex bedforms and a strong sulphate signature throughout the section¹³, indicating that sulphate-rich bedrock at least 500 m thick underlies the entire Meridiani unit. An enormous amount of precipitation would be required to account for this sulphate, yet there is no evidence for atmospheric precipitation of the required magnitude (for example, runoff channels). Furthermore, the area of sulphate mineralization appears to coincide with the outlines of the Meridiani unit, and it is unlikely that an atmospheric source would be spatially restricted to this unit.

It appears, therefore, that the most reasonable explanation for the exposed bedrocks is that they were originally emplaced as a basaltic pyroclastic deposit^{14–16} and subsequently altered through reaction with an aqueous sulphuric acid solution derived from condensation of SO_2 - and H_2O -rich vapours in a solfatar (fumarole)-like setting. In addition to high S abundances, the predominance of haematite and ferric sulphates^{2,6} indicates the rocks have been extensively oxidized relative to pristine martian basalts. Because ferrous Fe-bearing minerals including magnetite and pyrite are thermodynamically stable in the presence of sulphate, reaction with sulphuric acid solutions cannot account for this oxidation. Ferrous Fe in the Meridiani bedrock must therefore have been oxidized either by O_2 from the atmosphere or by reaction with water followed by escape of H_2 to space¹¹ (the current martian atmosphere contains $\sim 10^{-5}$ bar O_2 (ref. 17), and early Mars may have had similar levels). Because the haematite spherules appear to be late-stage concretions^{2–4} that required temperatures $>100^\circ\text{C}$ to form¹¹, oxidation probably occurred after the rocks had already been extensively altered but were still warm. Mineral precipitation from volcanic vapours can also account for the high Br:Cl ratios observed in the bedrocks, either through deposition from vapours enriched in Br or by fractionation of Br relative to Cl during precipitation. Enrichment in Br relative to Cl is frequently observed in volcanic gases^{18,19}, and mineral precipitates with high Br:Cl ratios are reported from volcanic fumaroles²⁰. Some of the sulphate minerals common in the Meridiani rocks probably represent precipitates left behind during later stages of alteration as the volcanic water evaporated away.

Although bedding features in the Meridiani rocks may be consistent with fluvial or aeolian deposition, other particle sedimentation processes might equally well explain their morphology. We suggest that volcanic deposition may account for the textures observed in Meridiani rocks, as morphologic bedding features including fine, sorted grains, planar bedding, low-angle cross-stratification (including festoon bedding) and ripple lamination are observed in base surge deposits of volcanic ash on Earth²¹ (Fig. 2). The lath-shaped vugs may be the remains of ferrous sulphates (or mixed Fe-Mg-Ca sulphates) formed in the early stages of alteration that became unstable during oxidation. Although minor relative to Mg- and Ca-sulphates, ferrous sulphate minerals such as szomolnokite and melanterite are reported from solfataras on Earth and may be more abundant in the Fe-rich rocks on Mars than in their terrestrial counterparts. These minerals are in the same crystal class as gypsum, and can form tabular crystals like those implied by the shape of the vugs. Given the extensive alteration of the rocks, most of the alteration probably occurred at high temperatures ($\sim 100^\circ\text{C}$ or higher), possibly soon after the volcanoclastic deposits were emplaced and still retained their original heat. At these elevated temperatures,

extended periods of time (for example, hundreds of thousands of years or more) would not have been required for completion of the hypothesized diagenetic sequence. Indeed, it seems feasible that at elevated temperatures the observed alteration could have occurred on timescales of a few thousand years or less.

Although the Meridiani unit is much larger than analogous terrestrial volcanic features, its size appears to be consistent with other martian volcanic deposits and with theoretical constraints. Geologic observations indicate that the equatorial regions of Mars are covered by globally distributed volcanic ash flows and/or air-fall deposits that are areally extensive ($\sim 3 \times 10^6 \text{ km}^2$) and more than a kilometre thick in some areas¹⁶, suggesting that very large scale explosive eruptions have occurred throughout the history of the planet. The Meridiani unit is comparable in size to pyroclastic deposits elsewhere on Mars, such as Hadriaca Patera²² ($1.1 \times 10^6 \text{ km}^2$), and an order of magnitude smaller than the Medusa Fossae deposits west of Tharsis that appear to be volcanic ash^{16,23}. The large scale of these deposits is consistent with theoretical considerations, which indicate that lower gravity and atmospheric drag on Mars would result in air-fall deposits and pyroclastic flows that are substantially larger than comparable deposits on Earth^{14,16,22,24}. Although the volcanic source for Meridiani is not immediately apparent, the sources of many terrestrial volcanoclastic deposits remain unknown because of subsequent massive burial by ash²⁵.

In contrast to the distribution of sulphates throughout the Meridiani unit, the haematite deposits appear to comprise only a very thin veneer capping the unit. Regional remote sensing¹² and local Opportunity^{1,2} data both imply that haematite exists as an unconsolidated, mobile lag deposit less than a few metres thick. Spectral data do not show a haematite signature in the deeply eroded sections of the light-toned outcrops that are stratigraphically below the materials observed by Opportunity, implying that the haematite spherules are not present throughout. Furthermore, lag deposits of haematite are not observed where the lower sections are eroded. These observations suggest that most of the Meridiani unit has been

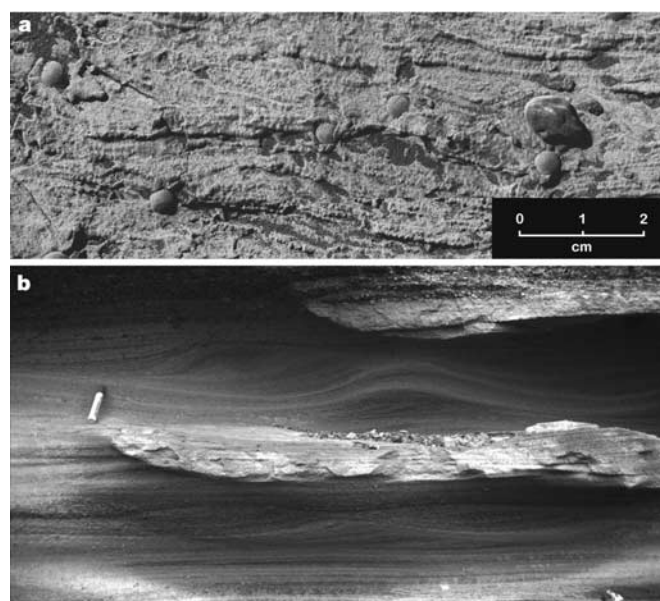


Figure 2 | Comparison of fine-scale morphologic features of bedrock from Meridiani Planum (a) with a terrestrial volcanoclastic deposit (b). Although the morphology of the martian rocks have been interpreted to be fluvial in origin^{1–6}, similar features are present in the volcanoclastic deposit, including planar bedding, festoon cross-bedding and low-angle cross-stratification. The volcanoclastic deposit is from the western district of Victoria, Australia. Pen at left of b is about 13 cm long. Image credits: NASA/JPL and MER team² (a) and G. Boxer (b).

pervasively altered by SO₂-bearing vapours, but that oxidation only occurred at the top of the unit, implying an atmospheric source.

To illustrate the feasibility of this scenario, we carried out thermodynamic calculations to model reaction of martian basalt with an acidic aqueous solution and SO₂ vapour (see Supplementary Information for details). These initial models are not intended to reproduce the exact temperature or alteration pathway of the Meridiani bedrock; rather, they are meant to be illustrative of the type of alteration processes that appear to be responsible for diagenesis of the Meridiani rocks. The model included two steps: reaction between martian basalt and a sulphuric acid solution at 100 °C with addition of SO₂ vapour, followed by oxidation of the rocks by addition of O₂. In the first step, fluid–rock reaction resulted in an alteration mineral assemblage composed of (in decreasing order of abundance): Fe- and Mg-rich smectite, Mg-rich saponite, anhydrite, pyrite, muscovite and quartz. Following oxidation in the second step, the alteration assemblage consisted of quartz, Mg-rich nontronite, haematite, anhydrite, alunite and diasporite. During both reaction steps, the fluid is dominated by free and complexed Mg²⁺, Na⁺, and SO₄²⁻ at mildly acidic pH (Supplementary Table S3). Evaporation of water from the fluid would result in additional precipitation of salts, primarily Mg-sulphates. Jarosite, a mineral observed in the Meridiani bedrocks, did not occur in any of our equilibrium models. Because jarosite is unstable in equilibrium with haematite, it apparently persists as a metastable mineral owing to kinetic constraints²⁶ and may have formed during evaporation or as a weathering product subsequent to other alteration.

The model proposed here for the Meridiani bedrocks has implications for the history of water and potential habitability of Mars that are significantly different from the sedimentary deposit/evaporite model proposed by the MER team^{1,2}. That model invokes a standing body of water persisting at the surface for an extended period of time, and implies that environmental conditions on early Mars were sufficiently clement during at least some periods for standing water to persist. In contrast, the model proposed here requires neither the prolonged presence of liquid water nor a climate substantially different from that of today. Although microbes are found in high-temperature, acidic solfatara environments on Earth^{27,28}, our model implies a shorter-lived and less benign environment for developing or sustaining life on Mars than implied by the sedimentary/evaporite model.

Received 28 February; accepted 28 October 2005.

- Squyres, S. W. *et al.* The Opportunity rover's Athena science investigation at Meridiani Planum, Mars. *Science* **306**, 1698–1703 (2004).
- Squyres, S. W. *et al.* *In situ* evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* **306**, 1709–1714 (2004).
- Herkenhoff, K. E. *et al.* Evidence from Opportunity's microscopic imager for water on Meridiani Planum. *Science* **306**, 1727–1730 (2004).
- Rieder, R. *et al.* Chemistry of rocks and soils at Meridiani Planum from the alpha-particle X-ray spectrometer. *Science* **306**, 1746–1749 (2004).
- Christensen, P. R. *et al.* Mineralogy at Meridiani Planum from the Mini-TES experiment on the Opportunity rover. *Science* **306**, 1733–1739 (2004).
- Klingelhöfer, G. *et al.* Jarosite and hematite at Meridiani Planum from Opportunity's Mössbauer spectrometer. *Science* **306**, 1740–1745 (2004).
- Fernández-Remolar, D. *et al.* The Tinto River, an extreme acidic environment under control of iron, as an analog of the Terra Meridiani hematite site of Mars. *Planet. Space Sci.* **52**, 239–248 (2004).
- Lodders, K. A survey of shergottite, nakhlite and chassigny meteorites whole-rock compositions. *Meteorit. Planet. Sci.* **33**, A183–A190 (1998).
- Laul, J. C. *et al.* Chemical systematics of the Shergotty meteorite and the composition of its parent body (Mars). *Geochim. Cosmochim. Acta* **50**, 909–926 (1986).
- Gellert, R. *et al.* Chemistry of rocks and soils in Gusev Crater from the alpha particle X-Ray spectrometer. *Science* **305**, 829–832 (2004).
- Catling, D. C. & Moore, J. M. The nature of coarse-grained crystalline hematite and its implications for the early environment of Mars. *Icarus* **165**, 277–300 (2003).
- Hynek, B. M. Implications for hydrologic processes on Mars from extensive bedrock outcrops throughout Terra Meridiani. *Nature* **431**, 156–159 (2004).
- Arvidson, R. E. *et al.* Spectral reflectance and morphologic correlations in eastern Terra Meridiani, Mars. *Science* **307**, 1591–1594 (2005).
- Hynek, B. M., Arvidson, R. E. & Phillips, R. J. Geologic setting and origin of Terra Meridiani hematite deposit on Mars. *J. Geophys. Res.* **107**, doi:10.1029/2002JE001891 (2002).
- Arvidson, R. E. *et al.* Mantled and exhumed terrains in Terra Meridiani, Mars. *J. Geophys. Res.* **108**, doi:10.1029/2002JE001982 (2003).
- Hynek, B. M., Phillips, R. J. & Arvidson, R. E. Explosive volcanism in the Tharsis region: Global evidence in the martian geologic record. *J. Geophys. Res.* **108**, doi:10.1029/2003JE002062 (2003).
- Owen, T. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 818–834 (Univ. Arizona Press, Tucson, 1992).
- Villemant, B. & Boudon, G. H₂O and halogen (F, Cl, Br) behaviour during shallow magma degassing processes. *Earth Planet. Sci. Lett.* **168**, 271–286 (1999).
- Wahrenberger, C. M. *Some Aspects of the Chemistry of Volcanic Gases* PhD thesis no. 12252, Swiss Federal Inst. Tech. Zurich (1997).
- Coradossi, N., Garavelli, A., Salamida, M. & Vurro, F. Evolution of Br/Cl ratios in fumarolic salammioniac from Vulcano (Aeolian islands, Italy). *Bull. Volcanol.* **58**, 310–316 (1996).
- Wohletz, K. H. & Sheridan, M. F. A model of pyroclastic surge. *Geol. Soc. Am. Spec. Pap.* **180**, 177–194 (1979).
- Crown, D. A. & Greeley, R. Volcanic geology of Hadriaca Patera and Eastern Hellas region of Mars. *J. Geophys. Res.* **98**, 3431–3451 (1993).
- Scott, D. H. & Tanaka, K. L. Ignimbrites of Amazonis Planitia region of Mars. *J. Geophys. Res.* **87**, 1179–1190 (1982).
- Wilson, L. & Head, J. W. Mars: Review and analysis of volcanic eruption theory and relationships to observed landforms. *Rev. Geophys.* **32**, 221–263 (1994).
- Ross, C. S., Smith, R. L. Ash-flow tuffs: Their origin, geologic relations, and identification. *US Geol. Surv. Prof. Pap.* **366**, 1–79 (1961); reprinted as *Spec. Publ.* 9 (New Mexico Geol. Soc., Washington DC, 1980).
- Elwood Madden, M. E., Bodnar, R. J. & Rimstidt, J. D. Jarosite as an indicator of water-limited chemical weathering on Mars. *Nature* **431**, 821–823 (2004).
- Huber, R., Huber, H. & Stetter, K. O. Towards the ecology of hyperthermophiles: biotopes, new isolation techniques, and novel metabolic properties. *FEMS Microbiol. Rev.* **24**, 615–623 (2000).
- Glamoclija, M., Garrel, L., Berthon, J. & López-García, P. Biosignatures and bacterial diversity in hydrothermal deposits of Solfatara Crater, Italy. *Geomicrobiol. J.* **21**, 529–541 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported in part by the NASA Astrobiology Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.M.M. (mccollom@lasp.colorado.edu).

LETTERS

Light echoes from ancient supernovae in the Large Magellanic Cloud

Armin Rest¹, Nicholas B. Suntzeff¹, Knut Olsen¹, Jose Luis Prieto², R. Chris Smith¹, Douglas L. Welch³, Andrew Becker⁴, Marcel Bergmann⁵, Alejandro Clocchiatti⁶, Kem Cook⁷, Arti Garg⁸, Mark Huber⁷, Gajus Miknaitis⁴, Dante Minniti⁶, Sergei Nikolaev⁷ & Christopher Stubbs⁸

The light from historical supernovae could in principle still be visible as scattered-light echoes centuries after the explosion^{1–6}. The detection of light echoes could allow us to pinpoint the supernova event both in position and age and, most importantly, permit the acquisition of spectra to determine the ‘type’ of the supernova centuries after the direct light from the explosion first reached Earth. Although echoes have been discovered around some nearby extragalactic supernovae^{7–13}, targeted searches have not found any echoes in the regions of historical Galactic supernovae^{14–16}. Here we report three faint variable-surface-brightness complexes with high apparent proper motions pointing back to three of the six smallest (and probably youngest) previously catalogued supernova remnants in the Large Magellanic Cloud, which are believed to have been thermonuclear (type Ia) supernovae¹⁷. Using the distance and apparent proper motions of these echo arcs, we estimate ages of 610 and 410 years for two of them.

As part of the SuperMACHO microlensing survey, we have been monitoring the central portion of the Large Magellanic Cloud (LMC) every other night for three months each year over the last four years (2001–4). Using an automated pipeline, we subtract point-spread-function matched template images from the recent epoch images. The resulting difference images are remarkably clean of the constant stellar background and are ideal for searching for variable objects.

The well-known echo of SN 1987A shown in Fig. 1 was straightforward to recover in the difference images with our pipeline. The high apparent motion of the echoes, often superluminal, allows simple detection in difference images. To search for very faint echoes, we have examined by eye all the variable objects discovered by our automatic pipeline. We found a number of very faint linear structures that had high apparent proper motions with vector directions inconsistent with the 1987A echo. For each structure, we estimated a vector direction, as shown in Fig. 2. Figure 3 shows the echo vectors extrapolated backward in time, pointing to the position of SN 1987A and three other well-defined positions as the origins of the echo complexes. The origins of the four echo complexes are listed in Table 1. The three unidentified echo origins correspond within arcminutes to the positions of known supernova remnants (SNRs)¹⁸ and also correspond to three of the six youngest SNRs¹⁷. These three SNRs are precisely the three that are classified as likely type Ia events based on the X-ray emission spectra.

Given the positional match with young SNRs and the high apparent proper motions of the variable diffuse light, we conclude that these newly detected structures are likely to be scattered light echoes from type Ia supernovae in the LMC. Planned spectroscopy of

the brightest knots in the three echo complexes should allow us to determine the types of the supernovae and confirm the classifications from the X-ray studies.

The theory of supernova light echoes (by which we mean the actual scattered light echo rather than fluorescence or dust re-radiation) predicts that light echoes can be seen even centuries after the first arrival of light from the explosion. Using a differential form of the equation for surface brightness (equation 7 in ref. 19), we find for two different supernovae, 1 and 2:

$$\Sigma_2 = \Sigma_1 + (V_{2_SN} - V_{1_SN}) - 2.5 \log_{10}[r_1 t_1 / (r_2 t_2)] \\ - 2.5 \log_{10}(\Phi_2 / \Phi_1)$$

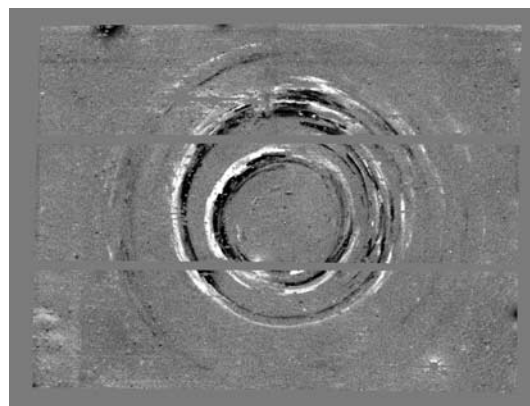


Figure 1 | The light echoes from SN 1987A. Data taken at the CTIO 4-m Blanco telescope with the MOSAIC imager in the VR filter were used to make this difference image; it shows epoch 2004.97 data minus epoch 2001.95 data, representing 17.8 and 14.8 yr after the explosion. Our SuperMACHO survey covers 24 degrees² in 68 pointings in an approximate rectangle 3.7° by 6.6° aligned with the LMC bar. The images are taken through our custom ‘VR’ filter (central wavelength $\lambda_c = 625$ nm, bandpass width $\Delta\lambda = 220$ nm) with exposure times of 60 s to 200 s, depending on the stellar densities. The field is 13.8′ by 18.4′, with north up and east left. Flux enhancements from 2004 are shown white and those from 2001 are shown black in this difference image. Faint echo arcs can be seen as far out as 6.6′ and 7.3′ from the explosion site, or 0.9 kpc and 1.1 kpc in front of SN 1987A. The VR surface brightness varies from 19.8 mag arcsec^{−2} to a limit of ~24 mag arcsec^{−2}, with one knot as bright as 19.3 mag arcsec^{−2}. The widths of the echoes are resolved, and are typically ~2.5″ across.

¹Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, La Serena, Chile. ²Department of Astronomy, Ohio State University, Columbus, Ohio 43210, USA. ³Department of Physics & Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada. ⁴Department of Astronomy, University of Washington, Seattle, Washington 98195, USA. ⁵Gemini Observatory, La Serena, Chile. ⁶Departamento de Astronomía y Astrofísica, Pontificia Universidad Católica de Chile, Santiago, Chile. ⁷Lawrence Livermore National Laboratory, Livermore, California 94550, USA. ⁸Department of Physics and Harvard-Smithsonian Center for Astrophysics, Harvard University, Cambridge, Massachusetts 02138, USA.

where Σ is the echo surface brightness, V_{SN} is the absolute supernova magnitude at maximum, r is the echo to supernova distance, t is the time between explosion and echo observation, and Φ is the Henyey-Greenstein phase function. Here we assume that the supernova light pulse duration is the same for the two supernovae, and that the composition, density and thickness of the dust sheets producing the echoes are identical. We also calculate the Φ function with forward scattering ($g = 0.6$, where g is the parameter measuring

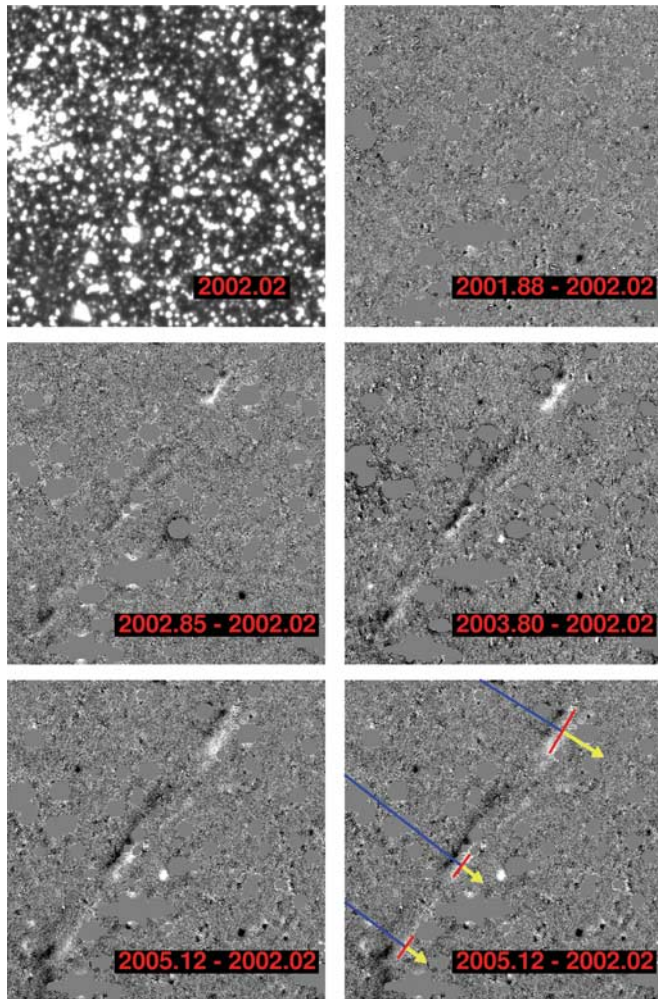


Figure 2 | Arcs of light echoes in the Large Magellanic Cloud from previously unseen supernovae. Top left, the unsubtracted (template) image, which includes the cluster Hodge 243. Top right, panel showing how cleanly the field subtracts with data taken 50 days earlier. Middle left and right, and bottom left, echo motion 1, 2 and 3 years after the template date. In these difference images (second epoch minus template, dates indicated in red), flux enhancements from the second epoch are shown white and flux enhancements from the template image are shown as black. Bottom right, the vector motions. Each echo is fitted with a straight line (red). The apparent proper motion is given by the yellow vector and extrapolated backwards (blue). The size of the yellow vector is proportional to the length of the echo segment fit. Saturated stars are masked out with grey circles. A number of faint variable stars appear as black or white spots. The vector was defined to be perpendicular to a linear fit to an echo segment, with the direction given by the apparent proper motion. Typical apparent proper motions range from $0.5'' \text{ yr}^{-1}$ to $2.4'' \text{ yr}^{-1}$ which, at the angular scale of the LMC of $0.77 \text{ light yr arcsec}^{-1}$, makes many of these structures have apparent superluminal velocities. The surface brightness ranges from $22.3 \text{ mag arcsec}^{-2}$ down to our limit of detection at $24 \text{ mag arcsec}^{-2}$. These echoes are located in echo complex no. 2, at right ascension (RA) 05 h 16 min 06 s, declination (dec.) $-69^\circ 17' 07''$, J2000. Each panel is $80''$ on a side with north up and east to the left.

the degree of forward scattering), and only include the angular terms. Scaling from the brightest echo knot of SN 1987A at $19.3 \text{ mag arcsec}^{-2}$, we find that a 500-yr-old type Ia supernova that exploded 150 pc behind a face-on dust sheet would produce a light echo with a surface brightness of $22.5 \text{ mag arcsec}^{-2}$ at an angular distance of 0.29° (250 pc radial distance from the supernova), assuming a type Ia supernova was 3.5 mag brighter than SN 1987A. At 1,000 years, the echo would be $24 \text{ mag arcsec}^{-2}$ at an angular distance of 0.5° or 420 pc from the explosion site. These surface brightness estimates are consistent with the echoes discovered here.

Supernova light echoes can be used to measure the structure and nature of the interstellar medium^{4,20,21} and, in principle, can be used to measure geometric distances²². The geometric relationship is $\rho = (ct(2z + ct))^{1/2}$, where ρ is the apparent projected radius of the light echo on the sky, z is the distance from the supernova to the dust sheet, and t is the time since peak brightness of the source. Given the known distance to the LMC and time of explosion, the echoes in Fig. 1 can be used to map out the structure of the dust²³.

What are the ages of the supernovae producing these echoes? A type Ia supernova in the LMC would reach an apparent V-band magnitude of $V \approx -0.5$, and would be the second or third brightest star in the southern sky for a few weeks. Lower limits on the supernova ages can be set from the absence of reported bright supernovae since the establishment of the Royal Observatory at the Cape of Good Hope in South Africa in 1820. An independent lower limit of 300 yr can be derived from the sizes of these SNRs assuming an unrealistic large constant shock velocity of $10,000 \text{ km s}^{-1}$.

We can use the apparent expansion velocity to crudely measure the ages of the supernova echoes. A simple differentiation of the formula above gives $v = c(z + ct)/\rho$, where v is the expansion velocity assuming the dust plane is perpendicular to the line of sight and c is the speed of light. Solving the two equations simultaneously, we find the age for echo 2 is $600 \pm 200 \text{ yr}$ with the dust $570 \pm 180 \text{ pc}$ in front of the supernova based on nine arcs, and for echo 3, an age of

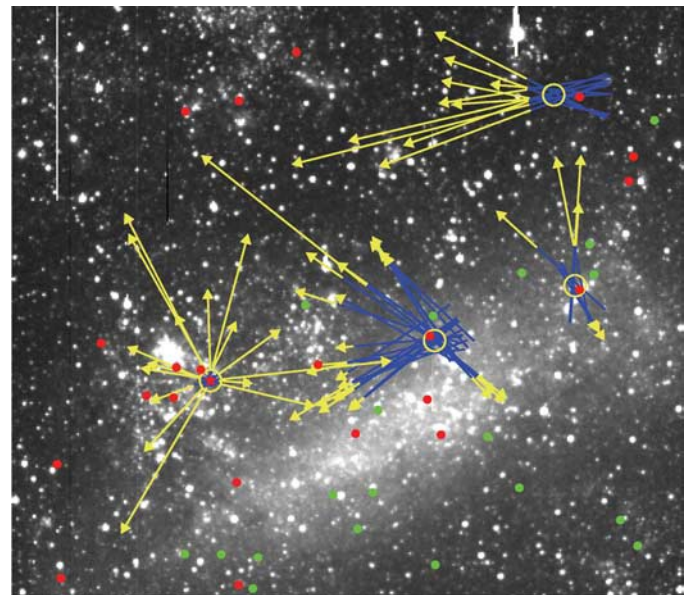


Figure 3 | A plot of the light echo vectors in the LMC. The vectors have the same meaning as in Fig. 2. The centres of the echo complexes are indicated by yellow circles. The lengths of the yellow vectors are $100\times$ the length of the echo arc. The source on the left marked with a star is SN 1987A. The green circles are the location of historical novae, and the red circles are the SNR locations²⁵. Evidently, the three unknown echo complexes point to three catalogued supernova remnants. We have estimated the position of the crossing point of the vectors by calculating the crossings of all pairs of vectors in each group, excluding any echo pair with a separation of less than $10''$.

Table 1 | Positions of supernova echo origins in the LMC

Echo complex	RA (h min s)	Dec. (° ')	Position error (arcmin)	δr (arcmin)	SNR name
1	05 35 30	−69 16	0.1	0.2	SN 1987A
2	05 19 14	−69 04	1	2.5	0519−69.0
3	05 11 17	−67 31	1	10.0	0509−67.5
4	05 09 19	−68 42	2	2.3	0509−68.7 (N103B)

Position errors are based on the intersection of the echo vectors. δr is the distance between the tabulated echo origin and the SNR. Coordinates (RA, right ascension; dec., declination) are equinox J2000. The error in the centroid was estimated from the averaged vector crossings.

400 ± 120 yr with the dust 340 ± 160 pc in front of the supernova based on six arcs. Echo 4 only had one arc with a superluminal velocity, giving an age of 860 yr. The alternative solutions to the equations gave ages greater than 2,500 yr, which are excluded on the basis of upper limits of less than 1,000 to 1,500 yr from the optical and X-ray observations²⁴. As a check on this technique, we measured an age for the SN 1987A echo of 15.9 ± 1.4 yr from 39 echo arc positions, which is consistent with the age of 1987A at the epoch of observation of 14.8 yr.

The uncertainties quoted above are the standard deviation of estimates from the different arcs. The uncertainties in the apparent proper motions, which are typically $0.1 \text{ arcsec yr}^{-1}$, propagate to age uncertainties of less than 50 yr. The largest uncertainty in the age estimates comes from the unknown inclinations of the dust sheets (assumed to be zero, or perpendicular to the line of sight). Allowing for inclinations leaves the upper limit on the ages unbounded, but lower limits can still be derived. If the dust sheets have inclinations of less than 60° , we find lower limits of 400 yr, 250 yr and 380 yr for the ages of echo 2, 3 and 4 respectively.

Also intriguing is the opportunity these light echoes provide for directly observing the spectral light from the historical supernovae themselves, as suggested²⁵ in 1940. Precise image subtraction techniques on nearby galaxies and in our own Galaxy with modern digital images can reach much fainter surface brightness limits than the early photographic surveys, and allow us to find echoes from supernovae as old as 1,000 yr or more. With the discovery of a bright echo knot, we might be able today to take a spectrum that represents the time average of the light at maximum of the Tycho, Kepler, SN 1006 or Cas A supernova. As an example, for a dust sheet 400 pc in front of the Tycho supernova with apparent magnitude $V_{\text{max}} = -6.5$, a distance of 3.1 kpc, and knots of densities similar to the highest density sheets near SN 1987A, the surface brightness would be $22 \text{ mag arcsec}^{-2}$. The arc would be at 6.5° from the Tycho SNR and would move at $30'' \text{ yr}^{-1}$. Scaling the typical echo width from the LMC, the Galactic echo would be $\sim 30''$ wide. A survey utilizing digital subtraction over an area of 100 degrees^2 could be able to recover these moving arcs.

Received 21 May; accepted 19 October 2005.

- Shklovskii, I. S. On the possibility of observations of light fronts from explosions of supernovae. *Astron. Circ. USSR* **306**, 2–3 (1964).

- van den Bergh, S. Light echoes from ancient supernovae. *Publ. Astron. Soc. Pacif.* **77**, 269–271 (1965).
- van den Bergh, S. The next Galactic supernova. *Astrophys. Space Sci.* **38**, 447–456 (1975).
- Chevalier, R. The scattered light echo of a supernova. *Astrophys. J.* **308**, 225–231 (1986).
- Schaefer, B. E. Light echoes: supernovae 1987A and 1986G. *Astrophys. J.* **323**, L47–L49 (1987).
- Patat, F. Reflections on reflexions—I. Light echoes in Type Ia supernovae. *Mon. Not. R. Astron. Soc.* **357**, 1161–1177 (2005).
- Crotts, A. Supernova 1987A in the Large Magellanic Cloud. *IAU Circ.*, 4561 (1988).
- Suntzeff, N. B. *et al.* The light echoes from SN1987A. *Nature* **334**, 135–138 (1988).
- Schmidt, B. P. *et al.* Reflections of past glory. *Astrophys. J.* **434**, L19–L23 (1994).
- Sparks, W. B. *et al.* Evolution of the light echo of SN 1991T. *Astrophys. J.* **523**, 585–592 (1999).
- Sugerman, B. E. & Crotts, A. P. S. Multiple light echoes from SN 1993J. *Astrophys. J.* **581**, L97–L100 (2002).
- Liu, J.-F., Bregman, J. N. & Seitzer, P. A scattered light echo around SN 1993J. *Astrophys. J.* **582**, 919–923 (2003).
- Cappellaro, E. *et al.* Detection of a light echo from SN 1998bu. *Astrophys. J.* **549**, L215–L218 (2001).
- van den Bergh, S. A preliminary search for light echoes from supernovae. *Astron. J.* **78**, 74–76 (1966).
- Boffi, F. R., Sparks, W. B. & Macchetto, F. D. A search for candidate light echoes: Photometry of supernova environments. *Astron. Astrophys. Suppl.* **138**, 253–266 (1999).
- Romaniello, M. *et al.* Very Large Telescope FORS1 imaging polarimetry of M83 (NGC 5236). I. Search for light echoes from historical supernovae. *Astrophys. J.* **629**, 250–258 (2005).
- Hughes, J. P. *et al.* ASCA observations of the Large Magellanic Cloud supernova remnant sample: Typing supernova from their remnants. *Astrophys. J.* **444**, L81–L84 (1995).
- Mathewson, D. S. *et al.* Supernova remnants in the Magellanic Clouds. III. *Astrophys. J. Suppl.* **58**, 197–200 (1985).
- Sugerman, B. E. K. Observability of scattered-light echoes around variable stars and cataclysmic events. *Astron. J.* **126**, 1939–1959 (2003).
- Couderc, P. Les auréoles lumineuses des novae. *Ann. Astrophys.* **2**, 271–302 (1939).
- Crotts, A. P. S. Discovery of optical echoes from supernova 1987A—new probes of the Large Magellanic Cloud. *Astrophys. J.* **333**, L51–L54 (1988).
- Sparks, W. B. A direct way to measure the distances of galaxies. *Astrophys. J.* **433**, 19–28 (1994).
- Xu, J., Crotts, A. P. S. & Kunkel, W. E. A Three-dimensional study using light echoes of the structure of the interstellar medium in front of SN 1987A. *Astrophys. J.* **451**, 806–815 (1995).
- Smith, R. C., Kirshner, R. P., Blair, W. P. & Winkler, P. F. Six Balmer-dominated supernova remnants. *Astrophys. J.* **375**, 652–662 (1991).
- Zwicky, F. Types of novae. *Rev. Mod. Phys.* **12**, 66–85 (1940).

Acknowledgements This Letter is based on observations obtained at NOAO, operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the NSF. C.S. thanks the National Science Foundation, the McDonnell Foundation, and Harvard University for their support of the SuperMACHO project. D.L.W. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC). The work of K.C., M.H. and S.N. was performed under the auspices of the US Department of Energy, National Nuclear Security Administration by the University of California, Lawrence Livermore National Laboratory. A.C. acknowledges support from FONDECYT. D.M. was partially supported by FONDAP. J.L.P. was funded by the OSU Astronomy Department Fellowship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.B.S. (nsuntzeff@ctio.noao.edu).

Extremely slow Drude relaxation of correlated electrons

Marc Scheffler¹, Martin Dressel¹, Martin Jourdan² & Hermann Adrian²

The electrical conduction of metals is governed by how freely mobile electrons can move throughout the material. This movement is hampered by scattering with other electrons, as well as with impurities or thermal excitations (phonons). Experimentally, the scattering processes of single electrons are not observed, but rather the overall response of all mobile charge carriers within a sample. The ensemble dynamics can be described by the relaxation rates, which express how fast the system approaches equilibrium after an external perturbation^{1–3}. Here we measure the frequency-dependent microwave conductivity of the heavy-fermion metal UPd₂Al₃ (ref. 4), finding that it is accurately described by the prediction for a single relaxation rate (the so-called Drude response⁵). This is notable, as UPd₂Al₃ has strong interactions among the electrons⁴ that might be expected to lead to more complex behaviour. Furthermore, the relaxation rate of just a few gigahertz is extremely low—this is several orders of magnitude below those of conventional metals (which are typically around 10 THz), and at least one order of magnitude lower than previous estimates for comparable metals. These observations are directly related to the high effective mass of the charge carriers in this material and reveal the dynamics of interacting electrons.

The concept of relaxation rates is the core of the theoretical description of the optical—that is, frequency-dependent—conductivity of metals. This goes back to the classical theory introduced by Drude in 1900⁵, but the relaxation-time approximation is used in the relevant quantum-mechanical treatments as well^{1–3}. Thus the complex frequency-dependent conductivity is expected to obey

$$\sigma_1 + i\sigma_2 = \sigma_0(1 - i\omega\tau)^{-1} \quad (1)$$

where σ_0 is the d.c. conductivity, $\omega = 2\pi f$ the angular frequency, and τ the relaxation time (the inverse of the relaxation rate). The real part σ_1 of the conductivity is characterized by a constant value, the d.c. conductivity, at low frequencies, and a decrease towards zero at the relaxation rate. The imaginary part σ_2 on the other hand exhibits a peak at the relaxation rate, with absolute value $\sigma_2(1/\tau) = \sigma_0/2$, and falls off at lower and higher frequencies. This simple frequency dependence, the so-called Drude response, only holds if a single relaxation time governs the charge dynamics of the complete system; deviations from Drude behaviour are expected for systems with strong interactions between the electrons.

The intermetallic compound UPd₂Al₃ belongs to the heavy-fermion materials⁶, which serve as prime examples of strongly correlated electron systems. Their electronic properties are dominated by the interaction of two subsets of electrons: one of (localized) *f* electrons, introduced by elements like Ce or U, and one of (delocalized) metallic band electrons. When a heavy-fermion compound is cooled to low temperatures, these two sets of electrons hybridize and create the so-called coherent state with a large density of states at the Fermi level. The relevant electrons now are mobile, but they react to external electric fields much more reluctantly than

conventional metallic electrons. Compared to free electrons, the effective mass m^* of the charge carriers is enhanced by up to a factor of 1,000, as reflected in the name ‘heavy fermions’. It leads to extraordinary values of several material properties—most famous are very high values of the electronic specific heat.

The mass enhancement m^*/m (where m is the mass of the non-interacting electron) goes hand in hand with a slowing-down of the characteristic timescales τ^*/τ : the relaxation rate $1/\tau^*$ of the heavy fermions is expected to be orders of magnitude lower than that of the non-interacting counterpart, $1/\tau$, and thus the response shifts to much lower frequencies. This effect has been addressed theoretically^{7,8} and experimentally⁹ since the 1980s. But the observation of the full electrodynamics remained an experimental challenge, and most optical studies barely support a quantitative determination of the relaxation rate¹⁰; the accessible frequency range is usually too high, and as a result—if at all—only the high-frequency tail of the response can be observed. But such assignments are problematic for several reasons; one that is particularly grave for heavy-fermion materials is the possible presence of additional excitations at low energies^{11,12}, possibly mimicking a Drude response if studied in too small a frequency range.

Apart from the shift of the relaxation rate towards very low frequencies, there is an additional reason why the optical properties of heavy fermions are of interest: if probed at very small energies (temperature and frequency), many heavy-fermion compounds, like UPd₂Al₃, are considered model systems for Fermi-liquid theory, the long-standing description of electron–electron interactions in metals^{2,13}. Electron–electron scattering is possible only very close to the Fermi surface, and this phase-space restriction leads to a distinct behaviour of the corresponding relaxation rate

$$1/\tau \propto (2\pi k_B T)^2 + (\hbar\omega)^2 \quad (2)$$

resulting in a quadratic d.c. resistivity— $\rho(T) = AT^2$, for instance. A relaxation rate like equation (2) causes characteristic deviations from the simple Drude response: whereas in equation (1)—describing the conductivity as a function of frequency—the relaxation rate $1/\tau$ was assumed constant, $1/\tau$ of equation (2) itself is a function of frequency. Combining the two equations to obtain the conductivity of the interacting electron system therefore leads to a more complicated frequency dependence. This effect has been addressed experimentally in optical studies on heavy fermions, but without conclusive results¹⁰.

UPd₂Al₃ is well characterized⁴, and has been studied previously with optical techniques¹⁴, but even investigations¹² down to 30 GHz could not detect the Drude response. As a consequence we study this material at yet lower frequencies with a novel broadband microwave spectrometer¹⁵: we have measured the complex reflection coefficient of epitaxial UPd₂Al₃ thin films^{16,17} at temperatures from 300 K down to 1.7 K in the frequency range 45 MHz to 40 GHz, and from these data we have directly calculated the complex conductivity.

The frequency and temperature dependence of the real part of the

¹Physikalisches Institut, Universität Stuttgart, D-70550 Stuttgart, Germany. ²Institut für Physik, Universität Mainz, D-55099 Mainz, Germany.

conductivity is shown as a surface plot in Fig. 1. In the low-frequency limit, the microwave data perfectly coincide with the d.c. conductivity, which is displayed on the left wall of the three-dimensional plot and was determined by a two-lead measurement simultaneously with the microwave measurements. The temperature-dependent d.c. transport is representative for heavy-fermion compounds⁴: upon cooling from room temperature, the conductivity first decreases, but below approximately 80 K the coherent state exhibits increasing conductivity with decreasing temperature like a normal metal. Around the Néel temperature $T_N = 14$ K there is a kink in the temperature dependence as this material becomes antiferromagnetic and spin-flip scattering is reduced. Thus the conductivity increases steeply and finally levels off at lower temperatures. Here we are more interested in the frequency dependence of the conductivity: whereas at high temperatures the conductivity is constant as expected for a conventional metal in this frequency range, the low-temperature data exhibits a very strong suppression towards higher frequencies—the Drude response we are interested in.

To study this roll-off quantitatively, a typical low-temperature spectrum is shown in Fig. 2. There real and imaginary parts of the conductivity are presented, together with a simultaneous fit of the two components following equation (1). Obviously this formula provides an excellent description of the frequency dependence, in particular around the relaxation rate, which is well within our accessible frequency range and apparent from the decrease in σ_1 and the coinciding maximum in σ_2 at 3 GHz. Thus the relaxation rate is unambiguously determined.

This relaxation rate is much lower than previously estimated for any other bulk metal, including the heavy-fermion materials. There have been several infrared studies on related compounds (for example, URu_2Si_2 (ref. 18), CeCoIn_5 (ref. 19), UPb_{13} (ref. 20), skutterudites²¹) where the relaxation rate could only be inferred from low-frequency extrapolation of the experimental data, and the characteristic roll-off was usually assumed to be in the range 30–300 GHz. Measurements right in that frequency range were

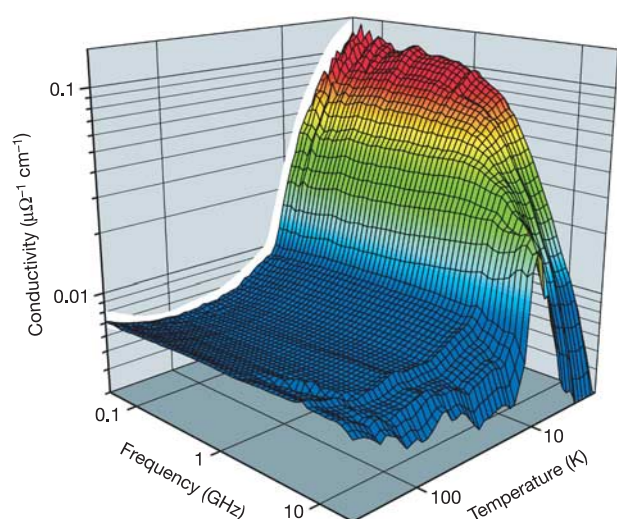


Figure 1 | Real part σ_1 of the complex conductivity of UPd_2Al_3 as a function of frequency (45 MHz to 20 GHz) and temperature (2 K to 300 K). The temperature dependence additionally plotted on the left wall (white line) represents the d.c. conductivity as determined with a two-lead measurement simultaneously with the microwave experiments. The coincidence with d.c. conductivity indicates that the low-frequency microwave data are obtained at frequencies $\omega \ll 1/\tau$. The strong high-frequency roll-off in $\sigma_1(\omega)$ for the lowest temperatures, on the other hand, occurs for $\omega > 1/\tau$. These two regimes also lead to the characteristic, non-monotonic temperature dependence at high frequencies with a maximum in σ_1 at the transition ($\omega = 1/\tau$) between these regimes.

limited to discrete frequencies of cavity resonators, and these showed that a strong decrease in σ_1 has to occur in the range 30–90 GHz for the heavy-fermion materials CePd_3 (ref. 22), CeAl_3 (ref. 23), UPt_3 (ref. 24) and U_2Zn_{17} (ref. 14). The low-temperature relaxation rate found in our experiment is at least an order of magnitude lower than inferred from those previous studies on heavy fermions, but our spectra extend continuously to frequencies well below the relaxation rate. This extremely low relaxation rate is particularly surprising, as the effective mass $m^*/m \approx 66$ is moderate by heavy-fermion standards⁴.

The only other electronic systems where comparable relaxation rates were deduced from experimental data are two-dimensional electron gases that are restricted to interfaces in semiconductor heterostructures²⁵. These structures can be grown with extremely high perfection, and therefore the mean free path ℓ of the electrons is very large (exceeding 100 μm for the best samples) and the scattering rate low. In contrast, in our system $\ell \approx 60$ nm is much shorter, and we observe the extremely low relaxation rates due to the low Fermi velocity of the electrons, approximately $1,300 \text{ m s}^{-1}$, that is three orders of magnitude smaller than in copper ($1.6 \times 10^6 \text{ m s}^{-1}$)¹.

Although this extremely slow relaxation is unprecedented in bulk metals, it is qualitatively understood within the framework of heavy fermions^{7,8}. But then it comes as a surprise that the frequency-dependent conductivity can be modelled extremely well within the simple Drude picture although our material has rather complex electronic properties governed by correlations. In particular, below 4 K our material exhibits a quadratic temperature dependence of the d.c. resistivity $\rho(T) = \rho_0 + AT^2$, a hallmark of a Fermi liquid. But in addition to electron–electron scattering as described by equation (2), we also have to consider the contribution ρ_0 due to imperfections of the sample, that is, scattering at impurities and the surface of our thin film. If the d.c. conductivity is governed by the quadratic Fermi-liquid temperature dependence, we expect the optical conductivity to exhibit the corresponding frequency dependence, that is, for the optical conductivity of equation (1) the relaxation time τ has to be the frequency-dependent τ of equation (2): but we do not observe this effect. The explanation is the impurity contribution ρ_0 , which is temperature and frequency independent and leads to the Drude behaviour: the influence of the Fermi-liquid quadratic frequency dependence (which can be estimated from the d.c. conductivity) is less than 1% and therefore negligible throughout our frequency range. Thus the shape of the spectrum presented in Fig. 2 is governed by impurity scattering, and can be described in the simple Drude picture.

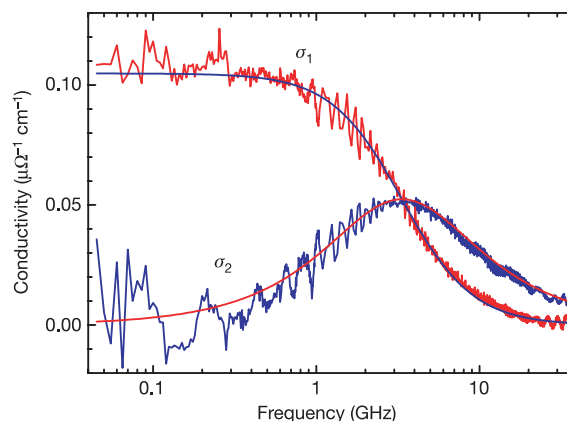


Figure 2 | Conductivity spectrum of UPd_2Al_3 at temperature 2.75 K; both real and imaginary parts (σ_1 and σ_2 , respectively) are shown. The fit ($\sigma_1 + i\sigma_2 = \sigma_0(1 - i\omega\tau)^{-1}$, $\sigma_0 = 0.105 \mu\Omega^{-1} \text{ cm}^{-1}$, $\tau = 4.8 \times 10^{-11} \text{ s}$) documents the excellent agreement of experimental data and the Drude prediction. The characteristic relaxation rate $1/\tau$ is marked by the decrease in σ_1 and the maximum in σ_2 around 3 GHz.

Our experiments show that even in a heavy-fermion compound where the characteristic frequencies are extremely low due to strong electronic correlations, Fermi-liquid effects might be negligible for the optical conductivity, revealing how difficult any experimental approach to optical signatures of Fermi-liquid behaviour will be. Heavy fermions are still prime candidates for the observation of such signatures, as the prefactor $A \propto (m^*)^2$ is orders of magnitude larger than found for normal metals. If there is any hope of finding a quadratic frequency dependence of the relaxation rate, one has to look at frequencies much higher than the relevant impurity scattering. For heavy fermions this means the study of ultra-pure samples in the frequency range addressed here, a combination that clearly demands considerable experimental improvements before any conclusive result is possible.

With the dominant role of impurity scattering at these rather moderate temperatures around 2 K (the d.c. resistivity of heavy-fermion compounds often display T^2 -behaviour below 1 K, with dominant impurity scattering only in the mK range) it is the more surprising that our experiments reveal the lowest relaxation rates among all heavy fermions. Nevertheless, strong electron–electron interactions are a prerequisite of these measurements, as they are responsible for the strong renormalization of the relaxation rate into this very low frequency range. To directly observe a ‘textbook’ Drude response over several decades in frequency as presented here, these low frequencies are essential. If the relaxation rate occurs at higher frequencies, as in the far-infrared for conventional metals, optical experiments encounter fundamental problems in obtaining phase-sensitive data on highly reflective samples, and furthermore the frequency range usually involves additional contributions to the optical conductivity, for example, interband transitions, that cannot be described by equation (1). Our system therefore represents a perfect mixture for exhibiting the Drude response: strong electronic interactions lead to an extremely slow relaxation suitable for direct spectroscopic detection, and the influence of impurities is strong enough to let the simple Drude description with a single relaxation rate be valid, but weak enough to let the relaxation rate remain in the observable low-frequency range.

METHODS

Sample preparation. To avoid the extremely high reflectivity of bulk samples, we study thin films. Thus we have grown high-quality epitaxial thin films of UPd_2Al_3 using molecular beam epitaxy techniques¹⁶: deposition onto $\text{LaAlO}_3(111)$ substrates by individually controlled coevaporation of the constituents U, Pd and Al. The physical properties of these thin films have been studied carefully and match those of bulk single crystals. The data presented here were obtained on a 150-nm-thick film.

Microwave measurements and data analysis. The microwave conductivity was determined with a Corbino spectrometer¹⁵: the reflection coefficient of the flat sample terminating a coaxial cable reveals the sample impedance. For cryogenic measurements (base temperature: 1.6 K) we have performed a full three-standards calibration that we could improve by using the superconducting state of the UPd_2Al_3 sample ($T_c = 2$ K) as a short. Using a superconducting short also allowed the very high upper-frequency limit of 40 GHz. Data analysis¹⁵ (from complex reflection coefficient via sample impedance to complex conductivity) is performed for each frequency separately, and requires no additional assumptions than that the sample is much thinner than the skin depth (which we have checked for the complete frequency and temperature range).

Microwave studies on metals at low temperatures usually have to consider the anomalous skin effect. There are two independent reasons why this is not the case here: first, our experiment directly probes the conductivity via the impedance of a thin film; that is, we do not have to consider the surface

impedance. Second, our material does not enter the anomalous regime as the Fermi velocity is very low, leading to a mean free path much shorter than the skin depth even for the extremely low scattering rate found here.

Received 25 July; accepted 12 September 2005.

1. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Saunders College Publishing, Fort Worth, 1976).
2. Ziman, J. M. *Principles of the Theory of Solids* (Cambridge Univ. Press, Cambridge, 1972).
3. Dressel, M. & Grüner, G. *Electrodynamics of Solids: Optical Properties of Electrons in Matter* (Cambridge Univ. Press, Cambridge, 2002).
4. Geibel, C. et al. Heavy-fermion superconductivity at $T_c = 2$ K in the antiferromagnet UPd_2Al_3 . *Z. Phys. B* **84**, 1–2 (1991).
5. Drude, P. Zur Ionentheorie der Metalle. *Phys. Z.* **1**, 161–165 (1900).
6. Grewe, N. & Steglich, F. in *Handbook on the Physics and Chemistry of Rare Earths* (eds Gschneidner, K. A. Jr & Eyring, L.) Vol. 14, 343–474 (Elsevier, Amsterdam, 1991).
7. Varma, C. M. Phenomenological aspects of heavy fermions. *Phys. Rev. Lett.* **55**, 2723–2726 (1985).
8. Millis, A. J. & Lee, P. A. Large-orbital-degeneracy expansion for the lattice Anderson model. *Phys. Rev. B* **35**, 3394–3414 (1987).
9. Webb, B. C., Sievers, A. J. & Mihalasin, T. Observation of an energy- and temperature-dependent carrier mass for mixed-valence CePd_3 . *Phys. Rev. Lett.* **57**, 1951–1954 (1986).
10. Degiorgi, L. The electrodynamic response of heavy-electron compounds. *Rev. Mod. Phys.* **71**, 687–734 (1999).
11. Donovan, S., Schwartz, A. & Grüner, G. Observation of an optical pseudogap in UPt_3 . *Phys. Rev. Lett.* **79**, 1401–1404 (1997).
12. Dressel, M. et al. Nature of heavy quasiparticles in magnetically ordered heavy fermions UPd_2Al_3 and UPt_3 . *Phys. Rev. Lett.* **88**, 186404 (2002).
13. Pines, D. & Nozières, P. *The Theory of Quantum Liquids* Vol. I, *Normal Fermi Liquids* (Benjamin, New York, 1966).
14. Degiorgi, L. et al. The electrodynamic response of heavy-electron materials with magnetic phase transitions. *Z. Phys. B* **102**, 367–380 (1997).
15. Scheffler, M. & Dressel, M. Broadband microwave spectroscopy in Corbino geometry for temperatures down to 1.7 K. *Rev. Sci. Instrum.* **76**, 074702 (2005).
16. Huth, M., Kaldowski, A., Hessert, J., Steinborn, T. & Adrian, H. Preparation and characterization of thin films of the heavy fermion superconductor UPd_2Al_3 . *Solid State Commun.* **87**, 1133–1136 (1993).
17. Jourdan, M., Huth, M. & Adrian, H. Superconductivity mediated by spin fluctuations in the heavy-fermion compound UPd_2Al_3 . *Nature* **398**, 47–49 (1999).
18. Bonn, D. A., Garrett, J. D. & Timusk, T. Far-infrared properties of URu_2Si_2 . *Phys. Rev. Lett.* **61**, 1305–1308 (1988).
19. Singley, E. J., Basov, D. N., Bauer, E. D. & Maple, M. B. Optical conductivity of the heavy fermion superconductor CeCoIn_5 . *Phys. Rev. B* **65**, 161101(R) (2002).
20. Brown Holden, A. A., Wardlaw, G. M., Reedyk, M. & Smith, J. L. Emergence of coherent transport in UBe_{13} : an optical investigation in both the normal and superconducting states. *Phys. Rev. Lett.* **91**, 136401 (2003).
21. Dordevic, S. V., Basov, D. N., Dilley, N. R., Bauer, E. D. & Maple, M. B. Hybridization gap in heavy-fermion compounds. *Phys. Rev. Lett.* **86**, 684–687 (2001).
22. Beyerlymann, W. P., Grüner, G., Dalichaouch, Y. & Maple, M. B. Relaxation-time enhancement in the heavy-fermion system CePd_3 . *Phys. Rev. Lett.* **60**, 216–219 (1988).
23. Awasthi, A. M., Degiorgi, L., Grüner, G., Dalichaouch, Y. & Maple, M. B. Complete optical spectrum of CeAl_3 . *Phys. Rev. B* **48**, 10692–10700 (1993).
24. Tran, P., Donovan, S. & Grüner, G. Charge excitation spectrum in UPt_3 . *Phys. Rev. B* **65**, 205102 (2002).
25. Burke, P. J., Spielman, I. B., Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. High frequency conductivity of the high-mobility two-dimensional electron gas. *Appl. Phys. Lett.* **76**, 745–747 (2000).

Acknowledgements The authors thank the Deutsche Forschungsgemeinschaft (DFG) for financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.S. (scheffl@pi1.physik.uni-stuttgart.de).

LETTERS

Global estimate of aerosol direct radiative forcing from satellite measurements

Nicolas Bellouin¹, Olivier Boucher¹, Jim Haywood¹ & M. Shekar Reddy²

Atmospheric aerosols cause scattering and absorption of incoming solar radiation. Additional anthropogenic aerosols released into the atmosphere thus exert a direct radiative forcing on the climate system¹. The degree of present-day aerosol forcing is estimated from global models that incorporate a representation of the aerosol cycles^{1–3}. Although the models are compared and validated against observations, these estimates remain uncertain. Previous satellite measurements of the direct effect of aerosols contained limited information about aerosol type, and were confined to oceans only^{4,5}. Here we use state-of-the-art satellite-based measurements of aerosols^{6–8} and surface wind speed⁹ to estimate the clear-sky direct radiative forcing for 2002, incorporating measurements over land and ocean. We use a Monte Carlo approach to account for uncertainties in aerosol measurements and in the algorithm used. Probability density functions obtained for the direct radiative forcing at the top of the atmosphere give a clear-sky, global, annual average of -1.9 W m^{-2} with standard deviation, $\pm 0.3 \text{ W m}^{-2}$. These results suggest that present-day direct radiative forcing is stronger than present model estimates, implying future atmospheric warming greater than is presently predicted, as aerosol emissions continue to decline¹⁰.

Anthropogenic biomass burning and industrial pollution aerosols are primarily emitted from agriculture and industry, although some biomass-burning emissions are due to naturally occurring large-scale fires in tropical and boreal regions. Natural aerosols are mostly mineral dust and marine aerosols, although some of the mineral dust emissions are due to human changes in land use. The aerosol direct radiative forcing (DRF) is defined as the perturbation of the radiative fluxes caused by anthropogenic aerosols (natural aerosols are not included). To obtain an estimate of the DRF, a measure of the anthropogenic aerosol loading and knowledge of their size distributions and refractive indices are needed. Remote-sensing measurements determine the total aerosol loading via the aerosol optical thickness (AOT), which is a measure of the wavelength-dependent aerosol extinction in the atmospheric column. Size distributions and refractive indices are determined from *in situ* measurements and used to compute the aerosol optical properties. The Moderate Resolution Imaging Spectrometer (MODIS) space instrument has operated onboard the Terra and Aqua platforms since December 1999 and May 2002, respectively. It provides the total AOT at $0.55 \mu\text{m}$ for clear-sky conditions over oceans and land surfaces^{6,7}, excluding deserts and snow-covered areas where the contribution from the surface to the measured signal is too large for accurate retrievals.

Determining the anthropogenic AOT from the total AOT requires additional information. The accumulation-mode fraction (AMF) is the fraction of the AOT from aerosols smaller than $1 \mu\text{m}$ in diameter (called accumulation-mode aerosols). The AMF is successfully retrieved by MODIS over oceans. Figure 1 presents airborne and

ground-based measurements of the AMF. Airborne measurements were made using Particle System Measurement probes by the Met Office C130 and FAAM BAe146 aircraft during the TARFOX and ADRIEX (industrial aerosols from North America and Italy, respectively), ACE-2 (pollution reaching the North Atlantic), SHADE (mineral dust from Sahara reaching Senegal and Capoverde), JET-2000 (mixture of mineral dust and biomass-burning in the gulf of Guinea), and SAFARI-2000 (biomass-burning in Namibia) field campaigns¹¹.

Sun-photometer measurements are from an analysis of several high-quality sites of the Aerosol Robotic Network (AERONET)¹². Natural aerosols alone are associated with AMFs smaller than 0.35 ± 0.05 . (From here on, the uncertain parameters are assumed to follow a gaussian distribution. Estimates of the mean value and the standard deviation are given.) Anthropogenic aerosols alone are associated with AMFs larger than 0.83 ± 0.05 . AMFs within these two boundaries are associated with mixtures of anthropogenic and natural aerosols. In such cases, the presence of a mixture of mineral dust and biomass-burning aerosols is implied by a significant TOMS (Total Ozone Mapping Spectrometer) aerosol index⁸, which includes only those aerosols that absorb in the ultraviolet, that is, mineral dust and biomass-burning aerosols. The AMF retrieved by MODIS is

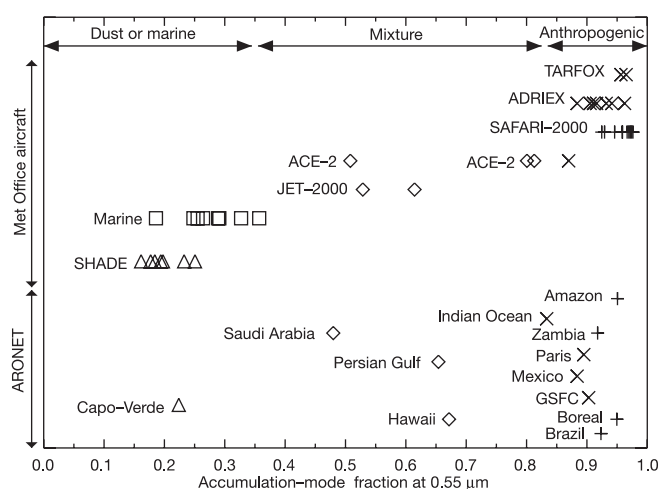


Figure 1 | *In situ* observations of the AMF. AMF at $0.55 \mu\text{m}$ as measured by the Met Office and FAAM aircraft during several field campaigns (top) and by the AERONET sun-photometers at selected sites¹². Measurements dominated by pollution aerosols are indicated by crosses, biomass-burning aerosols by plus signs, mixtures of marine and/or mineral dust and/or anthropogenic aerosols by diamonds, marine aerosols by squares and mineral dust by triangles. Airborne measurements of marine aerosols come from flights operated in clean atmospheric conditions.

¹Met Office, Exeter EX1 3PB, UK. ²NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey 08540, USA.

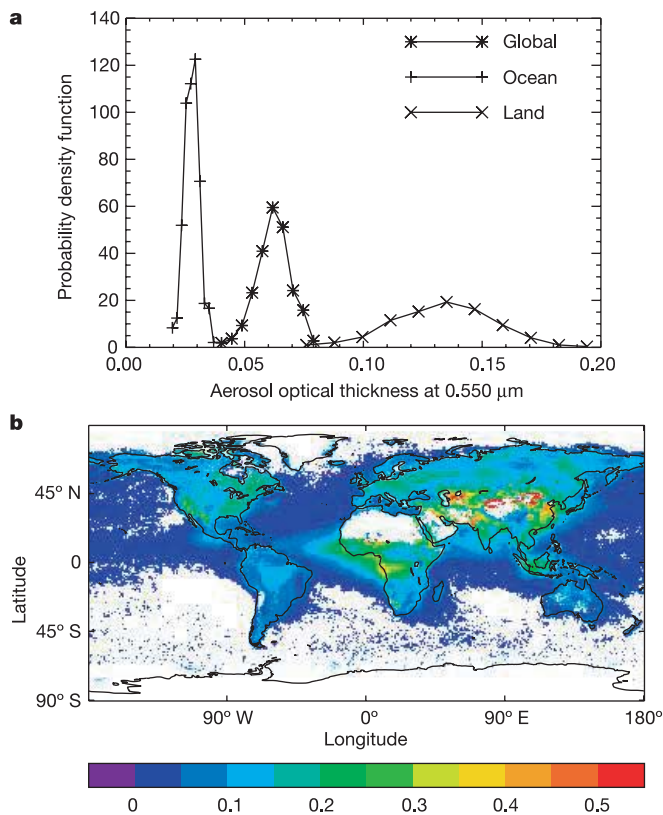


Figure 2 | Anthropogenic AOT at 0.55 μm. **a**, PDFs of the annual, global average over ocean, land and globally. **b**, Distribution for the year 2002. This is the mean of 250 experiments.

unfortunately not considered reliable over land surfaces. Five global models, including aerosol representations^{2,13–16} and using the same aerosol emissions, show that $47 \pm 9\%$ of the AOT over land is due to anthropogenic aerosols, on a global, annual average. To account for regional differences, we use different anthropogenic fractions over the six regions of our analysis (see Methods and Table 1). The standard deviation of the anthropogenic fraction is large for most of the regions, reflecting uncertainties in modelling aerosol transport and removal.

Uncertainties also exist in satellite products and algorithm parameters. For this reason, we do not give a single estimate of the anthropogenic AOT and DRF, but estimate probability density functions (PDFs) obtained from a Monte Carlo approach. Such a PDF is presented on Fig. 2a for the anthropogenic AOT from a set of 250 experiments. Over oceans, the relative accuracy of the MODIS aerosol algorithm and the robustness of the identification algorithm

translate into a narrow PDF, centred at an average AOT of 0.03 with a standard deviation of 0.003. The smaller accuracy of the MODIS AOT and the necessary use of models widen the PDF and the AOT over land is 0.13 ± 0.02 . On a global average, we estimate anthropogenic aerosols to have an AOT of 0.06 ± 0.01 . The annual distribution of the anthropogenic AOT, defined as the mean of our ensemble experiments and shown in Fig. 2b, illustrates the identification of the anthropogenic industrial pollution and biomass-burning aerosols achieved by the algorithm. Anthropogenic aerosols are shown to be significant contributors to the total AOT over oceans downwind of major biomass-burning events in central and southern Africa, and off the coasts of North America, Europe, China and India. Over land, largest AOTs are observed in the biomass-burning areas of Africa and South America, and the large industrial emissions from China and India are also apparent. We also note that the continuity between land and ocean is good, although not perfect.

The aerosol size distributions and refractive indices needed to convert the anthropogenic AOT to the DRF at the top of the atmosphere and at the surface are taken from AERONET measurements¹². A representative AERONET site is assigned to each of the above six regions (Table 1). The single-scattering albedo (SSA), defined as the ratio of scattering to extinction, is also shown. The most absorbing aerosols, corresponding to the smallest SSAs, are found in the biomass-burning regions and developing countries. The use of a single site to characterize a large region may seem unjustified: however, according to our analysis of AERONET sites, SSAs in the African and South American regions range from 0.85 to 0.91 and 0.89 to 0.94, respectively. The standard deviations used in the Monte Carlo approach encompass this regional variability as well as measurement errors. The PDFs for the DRF are presented in Fig. 3. The clear-sky DRF is $-1.9 \pm 0.3 \text{ W m}^{-2}$ at the top of the atmosphere, and $-4.4 \pm 0.6 \text{ W m}^{-2}$ at the bottom of the atmosphere, on a global average. The uncertainty in the DRF is due to the uncertainty in the anthropogenic AOT, but also to the uncertainty in the aerosol SSA. The largest DRFs are over land surfaces, implying that much work is needed to improve satellite retrievals over such surfaces. The difference between the top of the atmosphere and surface DRF corresponds to the energy absorbed in the aerosol layer, which amounts to 2.5 W m^{-2} for our best, clear-sky estimate.

We can approximate the all-sky DRF by assuming that the cloudy-sky contribution is negligible. Hence, the all-sky DRF is simply the clear-sky DRF multiplied by the clear-sky fraction (or one minus the cloud fraction). Using the MODIS cloud fraction¹⁷, the all-sky DRF is $-0.8 \pm 0.1 \text{ W m}^{-2}$ at the top of the atmosphere, and $-1.9 \pm 0.2 \text{ W m}^{-2}$ at the surface. In fact, the cloudy-sky contribution is likely to be either negligible for scattering aerosols or positive for absorbing aerosols above cloud¹⁸. The all-sky DRF is then certainly less negative than our estimate of -0.8 W m^{-2} , but this value cannot be improved until satellite observations supply the vertical profiles of aerosol and clouds.

The clear-sky DRF does not suffer from assumptions in the

Table 1 | Regional boxes used in the anthropogenic DRF estimation

	Boundaries	Anthropogenic fraction over land	AERONET site	SSA at 0.55 μm
North America	90° N–30° N 180° W–30° W	0.56 ± 0.21	GSFC (USA)	0.98 ± 0.02
Eurasia	90° N–30° N 30° W–180° E	0.54 ± 0.16	Creteil (France)	0.94 ± 0.03
Central America	30° N–0° 120° W–60° W	0.43 ± 0.11	Mexico City (Mexico)	0.90 ± 0.02
South America	30° N–90° S 180° W–30° W	0.35 ± 0.09	Brazil	0.91 ± 0.03
Africa, Oceania	30° N–90° S 30° W–180° E	0.43 ± 0.17	Mongu (Zambia)	0.86 ± 0.015
Indian Ocean	30° E–120° E 30° N–10° S	0.51 ± 0.15	Maldives	0.91 ± 0.03

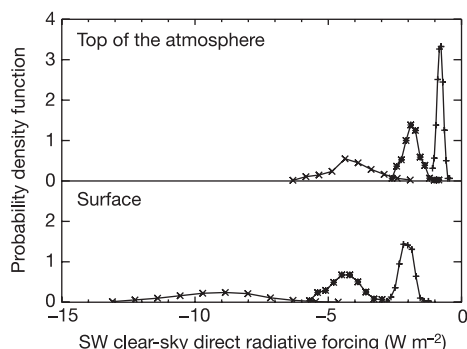


Figure 3 | PDFs of the clear-sky shortwave aerosol DRF on an annual, global average. Symbols are as used in Fig. 2a.

radiative effect of aerosols in cloudy sky, and is *prima facie* expected to compare well against the model estimates. However, there is a significant discrepancy between our clear-sky DRF (-1.9 W m^{-2}) and that from models (-0.5 to -0.9 W m^{-2}). Detailed comparisons with one such representative model revealed that the discrepancy stems from (in order of importance) too bright a surface albedo over both ocean and land in the model, too small an Ångström coefficient in the model, too small an optical thickness over land in the model, and to a smaller extent differences in global sampling, the radiative transfer code used to compute the radiative fluxes, the state of the mixture of the modelled aerosols and the vertical profiles of aerosol and water vapour. These results show that, although extreme negative values for the DRF are very unlikely, the DRF may be significantly stronger than current model estimates. Consequently, continued aerosol emission controls may lead to a stronger warming than current model predictions¹⁰.

METHODS

The PDFs are obtained from a set of 250 experiments. Each experiment derives an anthropogenic AOT and DRF. The algorithm uses daily data at the $1^\circ \times 1^\circ$ resolution from the MODIS products MOD08_D3. The MODIS AOT is corrected for the bias identified over both ocean and land surfaces¹⁹. Note that mineral dust and marine aerosol AOTs and DRFs are also estimated by the algorithm, but are left out of this study.

Anthropogenic AOTs and DRFs. Over ocean, the AMF retrieved by MODIS is reliable. Using the thresholds identified from Fig. 1, anthropogenic aerosols are immediately identified for AMFs larger than 0.83 ± 0.05 . The total AOT is nevertheless corrected for a marine aerosol background optical thickness. For AMFs ranging from 0.35 ± 0.05 to 0.83 ± 0.05 , grid-boxes with mixed mineral dust and biomass-burning aerosols are identified if the monthly averaged TOMS aerosol index is larger than 1.0 ± 0.15 . The anthropogenic aerosol receives the accumulation-mode part of the total AOT after correcting for the marine background AOT. The assumption that the MODIS AMF is entirely of anthropogenic origin may be inaccurate for dust outbreaks over ocean areas where the AMF is typically 0.5. Over the Atlantic Ocean west of the Sahara, from May to September, such misidentifications are estimated to overestimate the global, annual-averaged anthropogenic AOT by at most 5%. The marine background AOT, τ_{marine} , is estimated from the SSM/I surface wind speed⁹, w (in m s^{-1}), using $\tau_{\text{marine}} = (0.006 \pm 0.001)w + (0.060 \pm 0.005)$. This formula is taken from the *in situ* measurements summarized in ref. 20. The standard deviations are chosen to make the gaussian distribution encompass the whole range of measured slopes and intercepts.

Over land, the AMF is unfortunately unreliable and is replaced by the anthropogenic fraction estimated from modelled anthropogenic and total AOT. The five global models we used participated in the AEROCOM project, and were run with the same prescribed emissions under pre-industrial and present-day conditions. The obtained regional values (Table 1) of the anthropogenic fraction are used to convert the total MODIS AOT to an anthropogenic AOT.

Aerosol size distributions and refractive indices are also needed to compute shortwave radiative fluxes at the top of the atmosphere and at the surface. Here, we use sun-photometer measurements from selected sites of the Aerosol Robotic

Network (AERONET)¹². To account for the variability in aerosol properties, six regional boxes are defined and summarized in Table 1. Finally, we need to represent the surface albedo. Over oceans, the albedo depends on the solar zenith angle and the wavelength²¹ and is calculated for a wind speed of 7 m s^{-1} . Over land, it is computed from MODIS retrievals of the albedo for direct and diffuse radiation, with a different albedo in the visible and near-infrared spectra²². Surface albedo is adjusted for the aerosol effect on the distribution of downward radiation between the direct and diffuse fluxes. Radiative transfer calculations are performed using a discrete-ordinate solver²³, with 24 shortwave wavebands and 24 streams. Over ocean, the DRF is computed as the difference between radiative fluxes including the identified natural and anthropogenic aerosols and those including natural aerosols only. Over land, the DRF is computed as the difference between all aerosols and natural aerosols only, the former using the observed total optical depth, the latter using the total optical depth multiplied by one minus the anthropogenic fraction. The 24-hour-averaged DRF is computed by integrating the instantaneous radiative forcing over the solar zenith angles as a function of latitude and day of the year. Averages are weighted by the fraction of clear sky in a $1^\circ \times 1^\circ$ pixel (also termed pixel counts).

Monte Carlo approach. Most of the uncertainties in the uncertain algorithm parameters were previously given. We make one random choice constrained by our uncertainty assumptions for each global parameter (that is, threshold AMFs and aerosol regional SSA) and for each experiment. Similarly, we make multiple random choices for local parameters subject to measurement errors (that is, MODIS AOT, AMF and aerosol grid-box SSA) within each experiment. The MODIS total AOT at 0.55 μm , τ_{550} , has a published uncertainty of $\pm 0.03 \pm 0.05\tau_{550}$ over ocean, and $\pm 0.05 \pm 0.15\tau_{550}$ over land¹⁹. The MODIS AMF has a large uncertainty of ± 0.25 .

Received 1 August; accepted 17 October 2005.

1. Ramaswamy, V. *et al.* (ed.) *Climate Change 2001: The Scientific Basis. Contribution of WGI to the Third Assessment Report of the IPCC* (Cambridge Univ., Cambridge, 2001).
2. Reddy, M. S. *et al.* Estimates of global multi-component aerosol optical depth and direct radiative perturbation in the LMDZT General Circulation Model. *J. Geophys. Res.* **110**, doi:10.1029/2004JD004757 (2005).
3. Roberts, D. L. & Jones, A. Climate sensitivity to black carbon aerosol from fossil fuel combustion. *J. Geophys. Res.* **109**, doi:10.1029/2004JD004676 (2004).
4. Bellouin, N., Boucher, O., Tanré, D. & Dubovik, O. Aerosol absorption over the clear-sky oceans deduced from POLDER-1 and AERONET observations. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017121 (2003).
5. Christopher, S. A. & Zhang, J. Shortwave aerosol radiative forcing from MODIS and CERES observations over the oceans. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL014803 (2002).
6. Kaufman, Y. J. *et al.* Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *J. Geophys. Res.* **102**, 17051–17068 (1997).
7. Tanré, D., Kaufman, Y. J., Herman, M. & Mattoo, S. Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *J. Geophys. Res.* **102**, 16971–16988 (1997).
8. Herman, J. R. *et al.* Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data. *J. Geophys. Res.* **102**, 16911–16922 (1997).
9. Wentz, F. A well calibrated ocean algorithm for SSM/I. *J. Geophys. Res.* **102**, 8703–8718 (1997).
10. Andreae, M. O., Jones, C. D. & Cox, P. M. Strong present-day aerosol cooling implies a hot future. *Nature* **435**, 1187–1190 (2005).
11. Osborne, S. R. & Haywood, J. M. Aircraft observations of the physical and optical properties of major aerosol types. *Atmos. Res.* **73**, 173–201 (2005).
12. Dubovik, O. *et al.* Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci.* **59**, 590–608 (2002).
13. Takemura, T., Nozawa, T., Emori, S., Nakajima, T. Y. & Nakajima, T. Simulation of climate response to aerosol direct and indirect effects with aerosol transport-radiation model. *J. Geophys. Res.* **110**, doi:10.1029/2004JD005029 (2005).
14. Collins, W. D. *et al.* Simulation of aerosol distributions and radiative forcing for INDOEX: Regional climate impacts. *J. Geophys. Res.* **107**, doi:10.1029/2000JD000032 (2002).
15. Stier, P. *et al.* The aerosol-climate model ECHAM5-HAM. *Atmos. Chem. Phys.* **5**, 1125–1156 (2005).
16. Kirkevåg, A. & Iversen, T. Global direct radiative forcing by process-parameterized aerosol optical properties. *J. Geophys. Res.* **107**, doi:10.1029/2001JD000886 (2002).
17. Platnick, S. *et al.* The MODIS cloud products: algorithms and examples from Terra. *IEEE Trans. Geosci. Remote Sens.* **41**, 459–473 (2003).
18. Keil, A. & Haywood, J. M. Solar radiative forcing by biomass aerosol particles over marine clouds during SAFARI-2000. *J. Geophys. Res.* **108**, doi:10.1029/2002JD002315 (2003).

19. Remer, L. A. *et al.* The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* **62**, 947–973 (2005).
20. Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O. & Slutsker, I. Effect of wind speed on columnar aerosol optical properties at Midway Island. *J. Geophys. Res.* **108**, doi:10.1029/2003JD003879 (2003).
21. Cox, C. & Munk, W. Statistics of the sea surface derived from sun glitter. *J. Mar. Res.* **13**, 198–227 (1954).
22. Schaaf, C. B. *et al.* First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.* **83**, 135–148 (2002).
23. Key, J. R. & Schweiger, A. J. Tools for atmospheric radiative transfer: STREAMER and FLUXNET. *Comput. Geosci.* **24**, 443–451 (1998).

Acknowledgements The work by N.B., O.B. and J.H. was supported by the UK Department for Environment, Food and Rural Affairs under the Climate Prediction Programme. We thank B. Crouzille for helping with the processing of MODIS data. M. Schulz and the AEROCOM participants are thanked for their efforts and for letting us use their data.

Author Contributions All authors contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.B. (nicolas.bellouin@metoffice.gov.uk).

LETTERS

Anisotropy of Earth's D'' layer and stacking faults in the MgSiO₃ post-perovskite phase

Artem R. Oganov¹, Roman Martoňák², Alessandro Laio², Paolo Raiteri² & Michele Parrinello²

The post-perovskite phase of (Mg,Fe)SiO₃ is believed to be the main mineral phase of the Earth's lowermost mantle (the D'' layer). Its properties explain^{1–6} numerous geophysical observations associated with this layer—for example, the D'' discontinuity⁷, its topography⁸ and seismic anisotropy within the layer⁹. Here we use a novel simulation technique, first-principles metadynamics, to identify a family of low-energy polytypic stacking-fault structures intermediate between the perovskite and post-perovskite phases. Metadynamics trajectories identify plane sliding involving the formation of stacking faults as the most favourable pathway for the phase transition, and as a likely mechanism for plastic deformation of perovskite and post-perovskite. In particular, the predicted slip planes are {010} for perovskite (consistent with experiment^{10,11}) and {110} for post-perovskite (in contrast to the previously expected {010} slip planes^{1–4}). Dominant slip planes define the lattice preferred orientation and elastic anisotropy of the texture. The {110} slip planes in post-perovskite require a much smaller degree of lattice preferred orientation to explain geophysical observations of shear-wave anisotropy in the D'' layer.

The stability and properties of the post-perovskite (pPv) phase of (Mg,Fe)SiO₃ at conditions of the Earth's D'' layer are extensively used to explain seismic features of this layer^{1–4}, to understand the observed geochemical anomalies⁶ and global dynamics and evolution of the Earth^{5,6,12}. The initial finding of pPv^{1,2} was achieved with input from both experiment and theory. Here, starting from MgSiO₃ perovskite (Pv) and applying a new simulation technique¹³, we obtain the pPv structure purely from first principles. This shows the potential of this simulation methodology and provides new insight into the mineralogy and physics of the Earth's D'' layer.

We use the method proposed by Martoňák *et al.*^{13,14} and based on the ideas of metadynamics¹⁵. In this method, one introduces an order parameter—we use the lattice vectors matrix $h = (h_{11}, h_{22}, h_{33}, h_{12}, h_{13}, h_{23})$ chosen in the upper triangular form. This order parameter follows a discrete evolution:

$$h^{t+1} = h^t + \delta h \frac{\phi^t}{|\phi^t|} \quad (1)$$

where δh is a stepping parameter, and the driving force $\phi^t = -\frac{\partial G^t}{\partial h}$ is calculated from the history-dependent Gibbs potential $G^t(h)$ containing Gaussians added on top of the real free energy surface $G(h)$:

$$G^t(h) = G(h) + \sum_{t' < t} W e^{-\frac{|h - h^{t'}|^2}{2\delta h^2}} \quad (2)$$

where W is the height of the Gaussians. The derivative of the first term on the right-hand side of (2) is:

$$-\frac{\partial G}{\partial h_{ij}} = V[h^{-1}(p - P)]_{ji} \quad (3)$$

where p is the internal pressure tensor, and P is the external pressure.

Pressure tensors are calculated from constant-NVT (N , V and T are the number of particles, volume and temperature, respectively) molecular dynamics simulations; adding Gaussians (2) and evolving h -matrices as described above allows one to fill the free energy wells and move the system across the lowest barrier into the domain of another structure. Thus, one finds new crystal structures and structural transformation pathways, and although the latter will in general depend on the system size, precious suggestions can be inferred. To make the exploration of the free energy surface as complete as possible, it is useful to repeat simulations starting from each structure found.

We have performed classical (using a simplified interatomic potential¹⁶ with the DL_POLY code¹⁷) and *ab initio* (using the VASP code¹⁸) simulations. While the main results discussed here were obtained *ab initio*, classical simulations were used for initial exploration of the system and for testing conditions for *ab initio* simulations (system size, run length, δh and W parameters). The

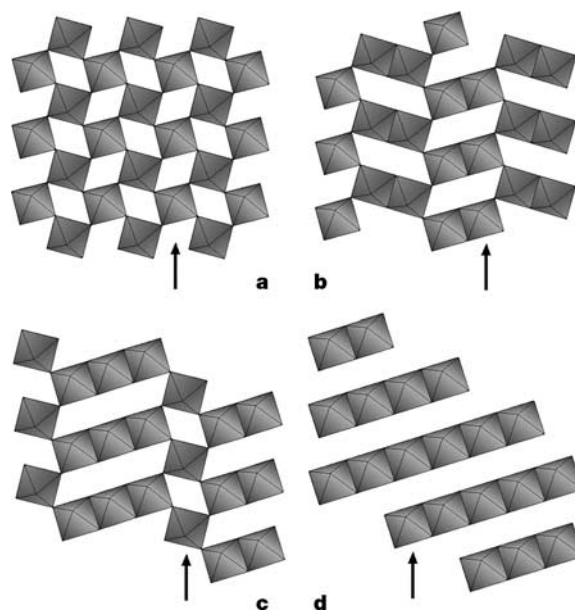


Figure 1 | MgSiO₃ polytypes found by metadynamics. **a**, Pv (space group *Pbnm*); **d**, pPv (*Cmcm*); **b**, **c**, newly found structures 2×2 (*Pbnm*) and 3×1 (*P2₁/m*), respectively. Only silicate octahedra are shown; Mg atoms are omitted for clarity. In the pPv structure, the previously expected slip plane is parallel to the sheets formed by silicate octahedra; the most likely slip plane identified here is shown by an arrow. Arrows also show slip planes in the other structures.

¹Laboratory of Crystallography, Department of Materials, ETH Zurich, HCI G 515, Wolfgang-Pauli-Strasse 10, CH-8093 Zurich, Switzerland. ²Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland.

Table 1 | Calculated equation-of-state parameters for MgSiO₃ polytypes

Phase	E_0 (eV)	V_0 (Å ³)	K_0 (GPa)	K'_0
Perovskite (<i>Pbnm</i>)	0	167.997	230.87	4.125
2 × 2 (<i>Pbnm</i>)	0.685	170.076	194.01	4.553
3 × 1 (<i>P2₁/m</i>)	0.645	169.563	198.55	4.515
Post-perovskite (<i>Cmcm</i>)	0.928	168.161	201.79	4.498

All results are given per 20 atoms. E_0 , V_0 , K_0 and K'_0 are the zero-pressure energy (relative to Pv), volume, bulk modulus and its pressure derivative, respectively.

ab initio simulations were based on the generalized gradient approximation¹⁹ and the all-electron projector augmented-wave (PAW) method^{20,21}. The time step for molecular dynamics was set to 1 fs, and in classical runs we used 4 ps for equilibration and 1 ps for calculating the pressure tensor; in *ab initio* calculations 0.7 ps was used for equilibration and 0.3 ps for pressure tensor calculations. Simulated conditions are 200 GPa, 2,000 K (classical) and 150 GPa, 1,500 K (*ab initio*).

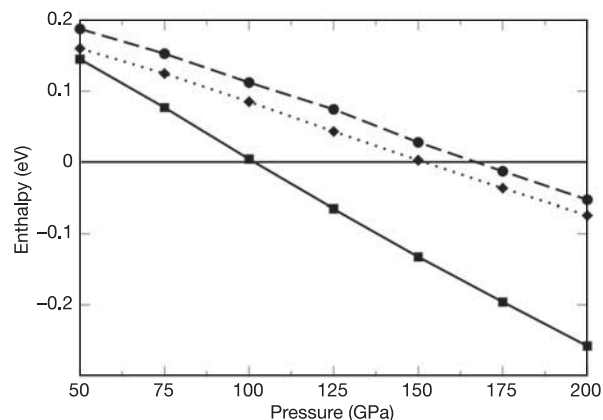
The supercells used in our calculations contained 160 atoms (4 × 1 × 2 for pPv, 2 × 2 × 2 for Pv), which is sufficiently big to encompass a large range of structures, while computationally tractable and providing clear transition paths. We used 450 eV plane-wave cut-off and the Γ -point for the Brillouin zone sampling; PAW potentials had [He] core (radius 1.52 atomic units, a.u.) for O, [Ne] core (1.5 a.u.) for Si, and [Ne] core (2.0 a.u.) for Mg. The metadynamics parameters we used are: $\delta h = 1$ Å and $W = 38$ kJ mol⁻¹ in the classical case, and 0.98 Å and $W = 32$ kJ mol⁻¹ in the *ab initio* case.

Starting from Pv (Fig. 1a), our *ab initio* simulations first found the 3 × 1 structure (Fig. 1c) and then the pPv structure (Fig. 1d). The reverse transition pathway was found in *ab initio* simulations starting from pPv. Classical simulations gave all these structures, plus the 2 × 2 structure (Fig. 1b). Table 1 reports the parameters of the Vinet equation of state²² fitted to our *ab initio* $E(V)$ results for these phases.

It is easy to see that these phases form a continuous family: by simple sliding of the {010}_{pV} planes of the Pv structure, we can generate all the other structures. Differing only in the stacking sequence of identical layers, these structures can be described as polytypes. Polytypes had been expected² in pPv since its discovery, because its structure contains layers of SiO₆-octahedra and polytypism is common in layered structures. However, the polytypes found here are radically different from those that were expected: they are not based on sheets of silicate octahedra parallel to {010}_{pPv}.

Figure 2 shows that all these polytypes become more favourable than Pv at sufficiently high pressure, but only the end members of this polytypic series, Pv and pPv, are thermodynamically stable at $T = 0$ K: Pv below 100 GPa, pPv above 100 GPa. Remarkably, the intermediate polytypes are only ~20–30 meV per atom higher in enthalpy around 100 GPa and could thus be easily stabilized by temperature and impurities and be present as minor phases in the D'' layer.

The stacking-fault enthalpy in pPv at 120 GPa is only 32 meV Å⁻² = 513 mJ m⁻², a small value similar to those found in metals at ambient pressure²³. It is common for polytypes that their typically low-energy stacking fault planes play the role of the dominant slip planes. The {110}_{pPv} slip planes at first seem counter-intuitive because they cut through the silicate sheets of the pPv structure, yet they are favourable. As shown by Legrand²³ on the example of hexagonal close-packed metals, the product of the stacking fault energy (or enthalpy) γ and the relevant shear elastic constant C_s determines the importance of a given slip plane. This criterion works very well (even though it does not explicitly account for dislocations), because it simultaneously accounts for ease of shear and formation of energetically favourable structures during plastic deformation. In our case, if the ratio $R = \frac{\gamma_{010}C_{s,010}}{\gamma_{110}C_{s,110}}$ is greater than 1, {110}_{pPv} slip planes should be preferred to {010}_{pPv} slip planes. By considering different types of {010}_{pPv} stacking faults

**Figure 2 | Enthalpies (relative to Pv, per formula unit) of MgSiO₃ polytypes as a function of pressure. Solid line, pPv; dashed line, 2 × 2 structure; dotted line, 3 × 1 structure.**

in post-perovskite, for the most stable ones (Supplementary Fig. 1) we found $\gamma_{010} = 330$ meV Å⁻². Using suitably transformed elastic constants² we obtain $R = 9.5$ at 120 GPa, ruling out {010}_{pPv} slip planes in favour of {110}_{pPv} slip.

Preferred orientation along {110}_{pPv} explains why in diamond-anvil cell experiments on pPv (for example, ref. 1) the {110} diffraction intensities are often vanishingly small. The fact that metadynamics could identify the most plausible slip plane in a single simulation is not surprising: the method by construction looks for the easiest non-elastic deformation mechanism and for the most energetically favourable structures along the deformation path. The mechanism of plastic deformation found here can operate also at pressures far away from the transition pressure and can be expected to be effective in analogous compounds. Indeed, using the analogy with CaTiO₃, Karato *et al.*¹⁰ concluded that the dominant slip plane in Pv should be {010}_{pV} (in the *Pbnm* setting used also in this paper); as seen in Fig. 1, sliding {010}_{pV} planes of Pv structure produces the pPv structure. The {010}_{pV} slip was also demonstrated to be important, though not dominant, in deformation experiments¹¹ conducted at 25 GPa.

These {110}_{pPv} slip planes in pPv call for a reinterpretation of the seismic anisotropy of the D'' layer. Using the method of ref. 24, we estimated seismic anisotropy of pPv texture with the {110}_{pPv} alignment; this required the elastic constants transformed to a new coordinate system: $C'_{ijkl} = \alpha_{ip}\alpha_{jq}\alpha_{kr}\alpha_{ls}C_{pqrs}$, where α is the transformation matrix and C_{pqrs} are the elastic constants² in the standard setting. Convective flow in the D'' layer is inclined in the regions of subduction, vertical in upwellings, and predominantly horizontal elsewhere. Because of the positive Clapeyron slope (refs 2, 25), the pPv layer will be thicker in cold subduction regions, so if anisotropy of the D'' layer is indeed related to pPv it should be more detectable in those regions. Detailed regional studies (see refs 8, 26, 27) indicate strong $\nu_{SH}/\nu_{SV} > 1$ anisotropy with an inclined axis in subduction regions, and variable anisotropy where the flow should be horizontal. By orienting the {110}_{pPv} slip planes (and the $\bar{1}\bar{1}0$ slip directions) vertically, we find that horizontally polarized shear waves propagate 4.1% faster than vertically polarized ones ($\nu_{SH}/\nu_{SV} = 1.041$). This anisotropy is larger than the previously reported value of 2.9% calculated with the assumption of horizontally located {010}_{pPv} slip planes. With the higher perfect-texture anisotropy obtained here, only 33% (65% in the regions of maximum anisotropy) alignment is required to reproduce the geophysically inferred anisotropy⁹. Orienting the slip directions at some angle (subduction angle) to the vertical would explain the inclined anisotropy invariably observed^{26,27} in the regions with $\nu_{SH}/\nu_{SV} > 1$; in the regions of horizontal convective flow $\nu_{SH}/\nu_{SV} < 1$ is expected. Previous interpretations based on {010}_{pPv} slip were unable to explain inclined

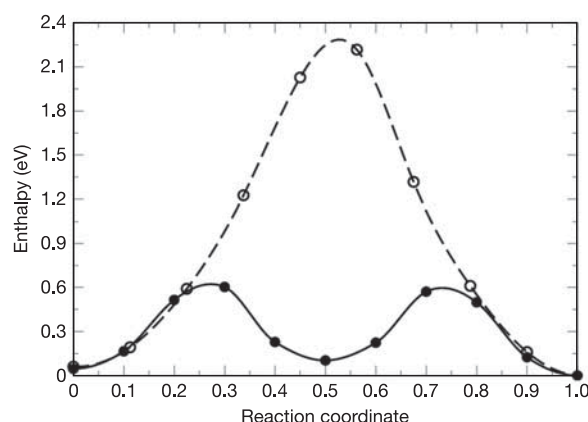


Figure 3 | Activation barrier for the Pv–pPv transition at 120 GPa. Dashed line, mechanism of ref. 25; solid line, mechanism proposed here. Enthalpies are given per formula unit. Along the reaction coordinate, the h -matrix smoothly changes from the values characteristic of Pv to those of pPv. At each point in the plot atomic positions were optimized under constraint of fixed h -matrix.

anisotropy and required unrealistically high degrees of lattice preferred orientation, 46% on average and 92% in maximally anisotropic regions.

Tsuchiya *et al.*²⁵ proposed a transition path from Pv to pPv based on shearing of the Pv structure in the $\{001\}_{\text{Pv}}$ plane. We observe this mechanism in simulations performed on a small 20-atom cell. For a larger, 160-atom system, however, we see a less cooperative mechanism with elements of nucleation: shear producing locally stacking faults with fragments of the pPv structure. First, we observe the transition from Pv to the 3×1 structure on the 15th metastep, and then to pPv on the 23rd metastep. Starting from pPv, we observe the reverse transition to perovskite following exactly the same pathway and again involving stacking faults. Using more degrees of freedom for atomic relaxation, the transition path obtained in a larger cell is by construction energetically more favourable. Direct calculation of the enthalpy as a function of the reaction coordinate (Fig. 3) shows that this effect is very large: instead of an enthalpy maximum in the middle of the transition path we have a local minimum corresponding to the intermediate 3×1 structure. As a consequence, the activation barrier at 120 GPa drops from ~ 2.3 eV for the pure-shear mechanism of ref. 25 to only 0.6 eV for our stacking-fault-mediated mechanism.

Our simulation technique—metadynamics—has enabled us to find the Pv–pPv transition mechanism and determine likely mechanisms of plastic deformation for both phases, involving the formation of stacking faults. Our predicted slip plane for Pv is consistent with experimental evidence. The predicted plastic slip of pPv is counterintuitive, but more consistent with geophysical observations than previous intuitive suggestions. In particular, it is now possible to explain the observed inclined character of anisotropy^{26,27}.

Received 29 July; accepted 15 November 2005.

1. Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. Post-perovskite phase transition in MgSiO_3 . *Science* **304**, 855–858 (2004).
2. Oganov, A. R. & Ono, S. Theoretical and experimental evidence for a post-perovskite phase of MgSiO_3 in Earth's D'' layer. *Nature* **430**, 445–448 (2004).
3. Iitaka, T., Hirose, K., Kawamura, K. & Murakami, M. The elasticity of the

MgSiO_3 post-perovskite phase in the Earth's lowermost mantle. *Nature* **430**, 442–445 (2004).

4. Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Elasticity of post-perovskite MgSiO_3 . *Geophys. Res. Lett.* **31**, L14603 (2004).
5. Hernlund, J. W., Thomas, C. & Tackley, P. J. A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* **434**, 882–886 (2005).
6. Ono, S. & Oganov, A. R. *In situ* observations of phase transition between perovskite and CaIrO_3 -type phase in MgSiO_3 and pyrolytic mantle composition. *Earth Planet. Sci. Lett.* **236**, 914–932 (2005).
7. Lay, T. & Helmberger, D. V. A shear velocity discontinuity in the lower mantle. *Geophys. Res. Lett.* **10**, 63–66 (1983).
8. Lay, T., Williams, Q. & Garnero, E. J. The core-mantle boundary layer and deep Earth dynamics. *Nature* **392**, 461–468 (1998).
9. Panning, M. & Romanowicz, B. Inferences on flow at the base of Earth's mantle based on seismic anisotropy. *Science* **303**, 351–353 (2004).
10. Karato, S., Zhang, S. Q. & Wenk, H. R. Superplasticity in Earth's lower mantle—evidence from seismic anisotropy and rock physics. *Science* **270**, 458–461 (1995).
11. Cordier, P., Ungar, T., Zsoldos, L. & Tichy, G. Dislocation creep in MgSiO_3 perovskite at conditions of the Earth's uppermost lower mantle. *Nature* **428**, 837–840 (2004).
12. Nakagawa, T. & Tackley, P. J. Effects of a perovskite-post perovskite phase change near core-mantle boundary in compressible mantle convection. *Geophys. Res. Lett.* **31**, L16611 (2004).
13. Martoňák, R., Laio, A. & Parrinello, M. Predicting crystal structures: The Parrinello-Rahman method revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).
14. Martoňák, R. *et al.* Simulation of structural phase transitions by metadynamics. *Z. Kristallogr.* **220**, 489–498 (2005).
15. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
16. Oganov, A. R., Brodholt, J. P. & Price, G. D. Comparative study of quasiharmonic lattice dynamics, molecular dynamics and Debye model in application to MgSiO_3 perovskite. *Phys. Earth Planet. Inter.* **122**, 277–288 (2000).
17. Smith, W., Todorov, I. T. & Leslie, M. The DL-POLY molecular dynamics package. *Z. Kristallogr.* **220**, 563–567 (2005).
18. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
19. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
20. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
21. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
22. Vinet, P., Rose, J. H., Ferrante, J. & Smith, J. R. Universal features of the equation of state of solids. *J. Phys. Condens. Matter* **1**, 1941–1963 (1989).
23. Legrand, B. Relations entre la structure électronique et la facilité de glissement dans les métaux hexagonaux compacts. *Phil. Mag.* **49**, 171–184 (1984).
24. Montagner, J.-P. & Nataf, H.-C. A simple method for inverting the azimuthal anisotropy of surface waves. *J. Geophys. Res.* **91**, 511–520 (1986).
25. Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Phase transition in MgSiO_3 perovskite in the earth's lower mantle. *Earth Planet. Sci. Lett.* **224**, 241–248 (2004).
26. Garnero, E. J., Maupin, V., Lay, T. & Fouch, M. J. Variable azimuthal anisotropy in Earth's lowermost mantle. *Science* **306**, 259–261 (2004).
27. Wookey, J., Kendall, J.-M. & Rumpker, G. Lowermost mantle anisotropy beneath the north Pacific from differential S-ScS splitting. *Geophys. J. Int.* **161**, 829–838 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Calculations were performed at ETH Zurich and CSCS (Manno). A.R.O. is grateful to P. Cordier, T. Ungar, G. Ferraris, T. Balic-Zunic, E. Makovicky and C. Thomas for discussions on various aspects of this work.

Author Contributions A.R.O. designed and performed this work and wrote the paper. Many ideas on plasticity and phase transformation mechanisms arose from discussions between A.R.O., R.M., A.L. and M.P.; R.M. and P.R. assisted A.R.O. in technical aspects of this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R.O. (a.oganov@mat.ethz.ch).

The pelvic fin and girdle of *Panderichthys* and the origin of tetrapod locomotion

Catherine A. Boisvert¹

One of the most marked transformations in the vertebrate transition to land was that of fins to limbs. This transformation involved not only the generation of morphological novelties (digits, sacrum) but also a shift in locomotory dominance from the pectoral to the pelvic appendage¹. Despite its importance, the transformation from pelvic fin to hindlimb is the least studied and least well-documented part of this transformation, which is bracketed by the osteolepiform *Eusthenopteron* and the early tetrapods *Ichthyostega* and *Acanthostega*, but is not directly illuminated by any intermediate form. *Panderichthys* is the closest tetrapod relative currently represented by complete fossils², but its pelvic fin skeleton has not been described. Here, I present the only known articulated pelvic fin endoskeleton and associated partial pelvis of *Panderichthys*. The pelvic girdle is even less tetrapod-like than that of the osteolepiform *Eusthenopteron*³, but the pelvic fin endoskeleton shares derived characteristics with basal tetrapods despite being more primitive than the pectoral fin of *Panderichthys*^{4,5}. The evolution of tetrapod locomotion appears to have passed through a stage of body-flexion propulsion, in which the pelvic fins played a relatively minor anchoring part, before the emergence of hindlimb-powered propulsion in the interval between *Panderichthys* and *Acanthostega*.

An articulated specimen of *Panderichthys rhombolepis* (Gross) from the Institute of Geology at Tallinn University of Technology, numbered GIT434-1 (previously Pi 1633), was collected by E. Mark-Kurik in 1972 at Lode quarry, Latvia. The quarry is part of the Middle Devonian Lode Formation, which is contemporary to the upper part of the Gauja Formation. The specimen, exposed in dorsal view, is preserved in three dimensions and is longitudinally complete from the snout to the pelvic region (Fig. 1a, b). Parts of the skull, anterior trunk and posterior trunk are missing from the left lateral portion because of mechanical excavator damage prior to collection (E. Mark-Kurik, personal communication), but enough remains of the cranial anatomy and squamation to allow a secure identification.

The right pelvic girdle and articulated fin are seen in internal view at the very end of the specimen (Fig. 1c). The fin has been rotated relative to the pelvis so that the flexor surface faces dorsomesially, probably crushing the articulation between femur and pelvis in doing so. The pelvic girdle itself is small: it measures 3.5 cm for a 90.5 cm long pre-pelvic body, corresponding to 3.86% of the body size as compared with 5% in *Eusthenopteron*³ and 7% in *Acanthostega* (measured from reconstructions⁶). The pelvic girdle is flat, club-shaped with a thickened and unfinished surface anteriorly; the posterior part has finished margins. There is no iliac ramus, and the pelvic fin articulates with the posterior end of the pelvis (Fig. 1d). Five fin elements are preserved in articulation, three axial elements (femur, fibula, fibulare) and two pre-axial radials (tibia, intermedium); all of these elements are very close to each other, suggesting that there was little articular cartilage between them in life (Fig. 1e).

The femur is very flat and broad with a narrow postaxial process overlapping the fibula. It narrows somewhat proximally towards the (poorly preserved) articulation with the pelvis. Three prominent longitudinal ridges, which were presumably for muscle attachment, traverse the femur. The dorsal-most ridge and a narrower median ridge begin mid-element and finish, respectively, at the end of the femur and at the fibula–fibulare articulation. The ventral-most ridge begins anteriorly to the other two and ends at the fibula–fibulare articulation. Because of their unique orientation and shapes, these ridges cannot easily be homologized to the very conserved muscle scars of basal tetrapods⁶ nor to either of the (slightly conflicting) femoral ridge patterns described for *Eusthenopteron*^{3,7}. The fibula is plate-like and much wider than the tibia, but is probably shorter. The proximo-dorsal portion of the fibula is covered by a thin layer of dermal scales, but these scales do not impede observation of the ridges and contours of the element. The tibia is rod-shaped, narrow and long. The fibulare and intermedium are covered by lepidotrichia, but their proximal margins show the same relative proportions as the fibula and tibia. A cross-section of the fibulare is visible under the lepidotrichia at the broken distal edge of the fin, showing that the endoskeleton continues under the lepidotrichial covering.

The morphology of the pelvic fin and girdle of *Panderichthys* is a combination of primitive, transitional and unique characteristics that provide crucial information about the evolution of the tetrapod hindlimb and the origins of tetrapod locomotion. During the transformation from the osteolepiform fins to tetrapod limbs, the appendages and girdles underwent a number of radical changes. The pelvic girdle became a weight-bearing structure by evolution of an ischium, a full mesio-ventral contact between the two sides of the girdle, an ilium, and a contact between the vertebral column and the girdle through a sacral rib⁸. Fore- and hindlimbs shifted laterally by reorientation of the glenoid and the acetabulum⁹. The pectoral girdle became detached from the skull by loss of the extrascapulars, posttemporal and supracleithrum, and became adapted for limb support and muscle insertion by enlargement of the scapulo-coracoid¹⁰. Lepidotrichia were lost and digits were gained¹. The proportions of the limb elements changed by elongation of the humerus⁵ and (more strongly) femur relative to the ulna+radius and fibula+tibia, and equalization of the lengths of the radius+ulna, and tibia+fibia, by shortening of the radius and tibia¹⁰. The postaxial processes of the ulnae¹¹ and the fibula were lost, and the radius and ulna, as well as the tibia and fibula, were realigned to be parallel^{3,6} rather than diverging⁸. In the course of this transition, there was a shift in locomotory dominance from the forelimb to the hindlimb, which was first demonstrated by *Acanthostega* and *Ichthyostega*¹.

In contrast to *Acanthostega*, in which the forelimb is slightly more primitive than the hindlimb (at least with regard to the relative lengths of the radius and ulna⁶), *Panderichthys* has a pelvic fin that is

¹Subdepartment of Evolutionary Organismal Biology, Department of Physiology and Developmental Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden.

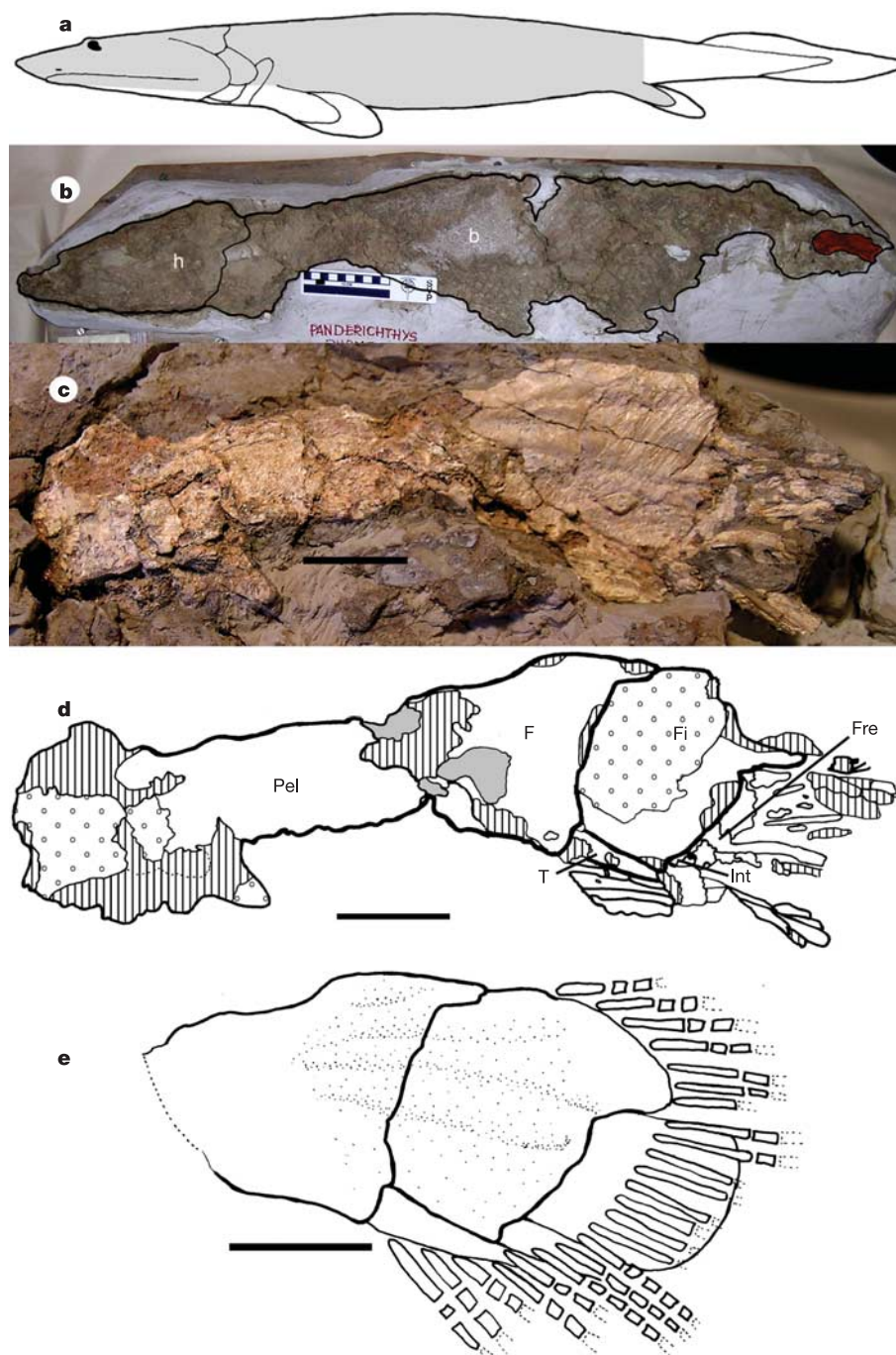


Figure 1 | Pictures and drawings of *Panderichthys rhombolepis*, specimen GIT434-1. **a**, Outline of the body of *Panderichthys*. Grey shading indicates preserved portions of *Panderichthys rhombolepis* specimen GIT 434-1. Redrawn from ref. 14. **b**, *Panderichthys rhombolepis* specimen GIT 434-1 with head (h) and body (b) outlined. The pelvic girdle and fin are shaded in orange. **c**, Pelvic girdle and fin. The matrix is distinguished from the fossil by an overlay of grey shading. **d**, Specimen drawing. F, femur; Fi, fibula; Fre,

fibulare; Int, intermedium (proximal end of the); Pel, pelvic girdle; T, tibia. Vertical hatching indicates broken bone; grey shading indicates matrix; circles indicate thin dermal bone covering. **e**, Reconstruction of the pelvic fin. Thick outline indicates preserved margin, thin outline indicates inferred margin, dotted lines indicate uncertain margin. Solid black scale bars, 10 mm.

more primitive than the pectoral (Fig. 2). The scapulocoracoid is intermediate in shape and size and the glenoid is oriented postolaterally⁹, whereas the pelvis of GIT434-1 lacks an ilium and retains a posteriorly oriented acetabulum (Fig. 1d). However, the pectoral and pelvic fins both share derived characteristics with tetrapod limbs, such as lack of a postaxial process on the ulna¹¹ and on the fibula (GIT434-1); parallel realignment of the ulna and radius¹¹, and the fibula and tibia (GIT434-1; Fig. 2c, d); and dorsoventral flattening of

the humerus and ulna⁵, and the tibia and fibula (GIT434-1). The pectoral fin is more tetrapod-like than the pelvic fin in that the shape of the humerus is intermediate between those of osteolepiforms and basal tetrapods⁵, whereas the femur of GIT-434-1 is more similar to that of osteolepiforms in that it retains the same length ratio to the fibula as in *Eusthenopteron*, and it lacks derived tetrapod characters such as an adductor blade and crest. Both the pectoral and pelvic fins of *Panderichthys* display very high ulnae to intermedium, fibula to

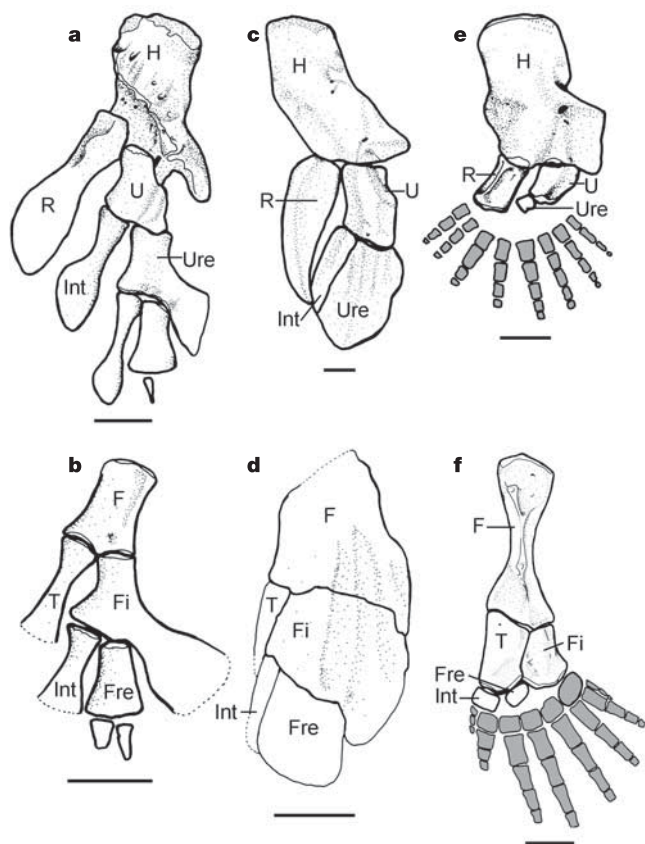


Figure 2 | Comparison of pectoral and pelvic fins. Pectoral (a, c, e) and pelvic fins (b, d, f) of *Eusthenopteron* (a, b), *Panderichthys* (c, d) and *Acanthostega* (e, f) all in ventral view. F, femur; Fi, fibula; Fre, fibulare; H, humerus; Int, intermedium; R, radius; T, tibia; U, ulna; Ure, ulnare. Thick outline indicates preserved margin; thin outline indicates inferred margin; dotted lines indicate uncertain margin. Scale bars, 10 mm. Panels a and b are redrawn from ref. 3, c is from ref. 11, scale bar is from ref. 5, and e and f are redrawn from ref. 6.

tibia, and fibulare to intermedium width ratios (the radial element in each case is much narrower than the axial element), characteristics that are not shared with either osteolepiforms or tetrapods (Fig. 2). Because of the great width of the axial elements, the articular surfaces are broad, but there seems to have been very little articular cartilage and thus little flexibility at the elbow⁵ and knee joint.

The disparity between the pectoral and pelvic fins and girdles of *Panderichthys* demonstrates that the transformations from fin to limb first began in the pectoral appendage and that most of the transformations in the pelvic appendage must have taken place between the *Panderichthys* and *Acanthostega* nodes of the tetrapod stem lineage. *Panderichthys* evidently was a 'front-wheel drive' animal that was incapable of tetrapod-like, hindlimb-propelled locomotion: the pelvic fin is much smaller than the pectoral fin; the pelvic girdle is not weight-bearing; the acetabulum is posteriorly oriented, rendering the tetrapod power-stroke impossible; and knee and elbow flexion would have been very limited. Nevertheless, both pectoral and pelvic fins appear to be adapted for substrate traction¹². This supports the hypothesis¹² that *Panderichthys* would have been capable of shallow-water or terrestrial locomotion driven by body flexion. The right pectoral fin would have functioned as an anchor while the trunk musculature would have bent the body, making it rotate around the fin. The right pelvic fin would then have been used as an anchor while the body would have been pushed forward,

returning to the starting position to repeat the cycle with the left side¹². The morphology of the pectoral girdle and fin^{4,5} also suggest that these elements would have played a more important role than the pelvic girdle and fin in lifting the front part of the animal off the ground during the first phase of movement. This mode of locomotion would have resembled closely that of the modern walking catfish *Clarias*¹².

Because the paired fin morphology of *Panderichthys* is defined substantially by a combination of primitive characters shared with osteolepiforms (mostly in the pelvic fin) and derived characters shared with tetrapods (many pectoral fin characteristics) rather than autapomorphies, at least part of the tetrapod stem lineage around the *Panderichthys* node must have displayed a combination of tetrapod-like pectoral fins with less limb-like pelvic fins. This suggests that the general locomotory pattern of *Panderichthys* characterized part of the tetrapod stem lineage between osteolepiforms and tetrapods. The evolution of tetrapod locomotion therefore seems to have passed through a 'front-wheel drive' stage powered by body undulation and pelvic fins as anchors, demonstrated by *Panderichthys*, before shifting to a 'rear-wheel drive' leg-powered walk in the interval between *Panderichthys* and *Acanthostega*. The absence of an iliac process raises questions about the supposed homology of the iliac ramus of *Eusthenopteron* and the ilium of tetrapods³. This absence, as well as the great disparity in the width between the bone pairs ulnare and intermedium, fibula and tibia, and fibulare and intermedium, might be autapomorphic or might represent aspects of a transitional morphology between osteolepiforms and tetrapods. In light of the unexpected transitional morphology of *Panderichthys*, data from more derived transitional forms such as *Elpistostege*¹³ are eagerly awaited.

Received 17 June; accepted 8 August 2005.

- Coates, M. I., Jeffery, J. E. & Ruta, M. Fins to limbs: what the fossils say. *Evol. Dev.* **4**, 390–401 (2002).
- Ahlberg, P. E. & Johanson, Z. Osteolepiforms and the ancestry of tetrapods. *Nature* **395**, 792–794 (1998).
- Andrews, M. S. & Westoll, S. T. The postcranial skeleton of *Eusthenopteron foordi* Whiteaves. *Trans. R. Soc. Edinb.* **68**, 207–328 (1970).
- Vorobyeva, E. I. The shoulder girdle of *Panderichthys rhombolpis* (Gross) (Crossopterygii), Upper Devonian, Latvia. *Geobios* **19**, 285–288 (1995).
- Vorobyeva, E. I. Morphology of the humerus in the *Rhipidistia* crossopterygii and the origin of tetrapods. *Paleontol. J.* **34**, 632–641 (2000).
- Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal interrelationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb.* **87**, 363–421 (1996).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* (Academic, London, 1980).
- Carroll, R. L., Irwin, J. & Green, D. M. Thermal physiology and the origin of terrestriality in vertebrates. *Zool. J. Linn. Soc.* **143**, 345–358 (2005).
- Vorobyeva, E. I. Evolution of the rhipidistia (Crossopterygii) pectoral girdle in the context of the origin of tetrapods. *Paleontol. Zh.* **3**, 3–11 (1998).
- Clack, J. A. in *Amphibian Biology* Vol. 4 *Paleontology: the Evolutionary History of Amphibians* (eds Heatwole, H. & Carroll, R. L.) 979–1029 (Surrey Beatty and Sons, Chipping Norton, Australia, 2000).
- Vorobyeva, E. I. The role of development and function in formation of tetrapod-like pectoral fins. *Zh. Obshch. Biol.* **53**, 149–158 (1992).
- Vorobyeva, E. I. & Kuznetsov, A. in *Fossil Fishes as Living Animals* (ed. Mark-Kurik, E.) 131–140 (Academy of Sciences of Estonia, Tallinn, 1992).
- Schultze, H.-P. & Arsenault, M. The *Panderichthys* fish *Elpistostege*: a close relative of tetrapods? *Palaeontology* **28**, 293–309 (1985).
- Vorobyeva, E. I. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ. Press, Ithaca, 1991).

Acknowledgements I thank E. Mark-Kurik for collecting and allowing access to the specimen; P. E. Ahlberg for discussions; and Vetenskapsrådet for financial support through a grant to P. E. Ahlberg.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to C.A.B. (catherine.boisvert@ebc.uu.se).

LETTERS

Dance reveals symmetry especially in young men

William M. Brown¹, Lee Cronk¹, Keith Grochow², Amy Jacobson¹, C. Karen Liu², Zoran Popović²
& Robert Trivers¹

Dance is believed to be important in the courtship of a variety of species, including humans, but nothing is known about what dance reveals about the underlying phenotypic—or genotypic—quality of the dancer^{1–6}. One measure of quality in evolutionary studies is the degree of bodily symmetry (fluctuating asymmetry, FA), because it measures developmental stability^{7,8}. Does dance quality reveal FA to the observer and is the effect stronger for male dancers than female? To answer these questions, we chose a population that has been measured twice for FA since 1996 (ref. 9) in a society (Jamaican) in which dancing is important in the lives of both sexes. Motion-capture cameras created controlled stimuli (in the form of videos) that isolated dance movements from all other aspects of visual appearance (including FA), and the same population evaluated these videos for dancing ability. Here we report that there are strong positive associations between symmetry and dancing ability, and these associations were stronger in men than in women. In addition, women rate dances by symmetrical men relatively more positively than do men, and more-symmetrical men value symmetry in women dancers more than do less-symmetrical men. In summary, dance in Jamaica seems to show evidence of sexual selection and to reveal important information about the dancer.

Darwin was the first to suggest that dance is a sexually selected courtship signal¹. If so, it should reveal genetic or phenotypic quality of the dancer. One such indicator of quality is degree of fluctuating asymmetry (FA), because it is inversely correlated with degree of developmental stability, which is an organism's ability to reach an adaptive end point despite ontogenetic perturbations^{7,8,10,11}. Across diverse taxa, increased FA is associated with increased morbidity, mortality, poor fecundity and other variables linked to natural and sexual selection^{7,8}. Most germane to the hypothesis that dance reveals underlying developmental stability is evidence that reduced fluctuating asymmetry is associated with locomotory traits or their functional effectiveness in several species, including humans^{12–19}. Likewise, bodily FA is inversely associated with attractiveness based on a person's odour²⁰, voice²¹ and facial appearance²². (Note that associations between FA and measures of sexual selection may sometimes be overestimated owing to publication bias and problems associated with small sample size²³.)

There are no studies in humans (or any other species) linking variation in dance quality with genetic and/or phenotypic quality. One difficulty is that dance evaluations are potentially confounded by the appearance of the dancer—factors such as clothing, culture, physical attractiveness and FA itself. To control for these potential confounds, we used motion-capture technology commonly used in medical and sport science to extract accurately complex movement²⁴, including dance²⁵ (see Supplementary Video 1 for how stimuli were constructed). Thus we were able to separate the phenotype of dance from the phenotype of the dancer. Motion-capturing of 183 human dancers was conducted in 2004 in Southfield, Jamaica. Each

individual danced alone to the same song (popular at the time in Jamaican youth culture) in the same place in front of the same film crew for one minute.

Forty dance animations were chosen on the basis of the level of fluctuating asymmetry of the dancer, using two measures of FA over time (1996 and 2002). Specifically, individuals in the top third on both FA measures were categorized 'asymmetrical' ($n = 20$), while individuals in the bottom third of both samples were categorized 'symmetrical' ($n = 20$). This experimental strategy helped to control for longitudinal changes in FA due to the accumulation of developmental errors, compensatory growth²⁶ or measurement error. Traits used to calculate composite relative FA were the elbow, wrist, knee, ankle, foot, third digit, fourth digit, fifth digit and ears. These traits were measured because they reveal true FA rather than directional biases, and have proven useful in numerous past studies of FA in humans^{9,20–22}. Dance animations of symmetrical (Supplementary Video 2) and asymmetrical (Supplementary Video 3) individuals were presented randomly to a sample of 155 Jamaican peers (themselves characterized for FA) for evaluation on a dance rating scale. No one interviewed afterwards ($n = 20$) was able to recognize any individual from his or her motion-capture video. Indeed, correct sex recognition in the current sample was only $62\% \pm 0.11$ (mean $\pm$ s.d.), with female evaluators ($64\% \pm 0.08$) being somewhat better at detecting the sex of the dancer than male evaluators ($60\% \pm 0.13$) (Mann–Whitney test, $Z = 2.25$, $P < 0.03$). Cronbach's alpha indicated that inter-rater agreement for the dance rating scale was 0.91 for the entire sample and for both subsamples of correct versus incorrect sex detections. Finally, there were no effects of correct sex identifications on the dance evaluations in the current model (see the Supplementary Background and Analyses). We conclude that evaluators based their judgments on motion-captured movements and not on recognition of the identity or sex of the dancer.

In species where fathers invest less than do mothers in their offspring, females are expected to be more selective in mate choice, and males to invest more in courtship display²⁷. Thus, we predicted that degree of symmetry would more strongly correlate with male dance ability, and females would be better discriminators. Assuming lower quality individuals (higher FA) are less attractive to the opposite sex, there may be selection for such individuals to shift their preferences downward towards individuals more likely to accept them as mates²⁸. Therefore, we tested whether greater FA of evaluators was associated with weaker preferences for symmetrical dancers.

As predicted, there was a significant effect of symmetry ($F_{1,34} = 16.34$, $P < 0.001$) and sex ($F_{1,34} = 10.99$, $P < 0.005$), and there was a significant interaction between them ($F_{1,34} = 4.46$, $P < 0.042$) on dance ability (corrected for the body mass index (BMI) and age of the dancer). Symmetrical males were evaluated as significantly better dancers (mean $\pm$ s.d. = 57.31 ± 10.65) than asymmetrical males (39.22 ± 9.23), accounting for 48% of the

¹Center for Human Evolutionary Studies, Department of Anthropology, Rutgers University, 131 George Street, New Brunswick, New Jersey 08901-1414, USA. ²Department of Computer Science and Engineering, University of Washington, Box 352350 Seattle, Washington 98195-2350, USA.

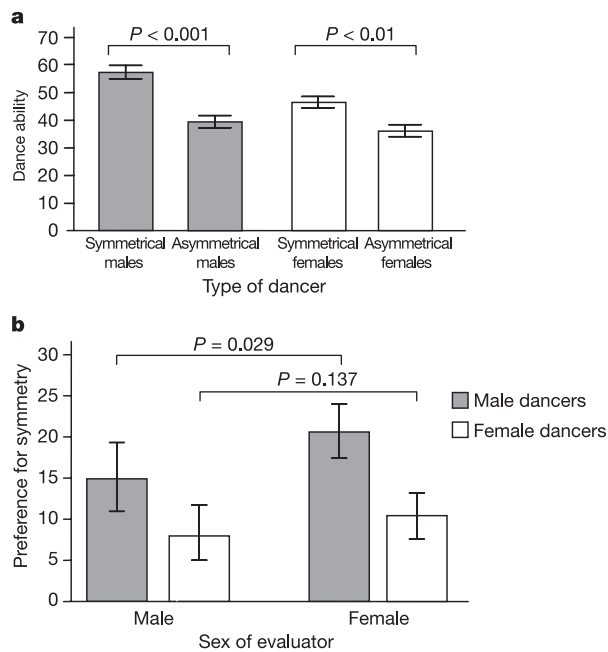


Figure 1 | Symmetry, dance ability and sex differences in evaluators' preferences for symmetry. **a**, Mean dance ability of males (filled bars throughout) and females (open bars throughout) by level of bodily symmetry. Error bars represent 95% confidence intervals. P values for within-sex are shown. **b**, Mean strength of symmetry preference (that is, evaluations of asymmetrical dancers subtracted from evaluations of symmetrical dancers) by sex of dancer and sex of evaluator. Error bars represent 95% confidence intervals. P value at top of panel indicates significantly greater female than male evaluator preference for symmetrical male dancers.

variance in dance ability ($t_{18} = 4.06$, $P < 0.001$) (Fig. 1a). Even though symmetrical females were significantly better dancers (45.53 ± 9.47) than asymmetrical females (35.58 ± 9.70), female symmetry only accounted for 23% of dance ability ($t_{18} = 2.32$, $P < 0.01$). Symmetrical males were significantly better dancers than asymmetrical females ($t_{18} = 3.21$, $P < 0.005$), but the difference in dance quality between asymmetrical male and female dancers was not significant ($t_{18} = 0.79$, $P > 0.45$). (The effect sizes in this study may be overestimates owing to the extreme group design where dancers were prescreened for degree of asymmetry, but this bias is not expected to affect the two sexes differently.) The dancers ranged in age from 14 to 19 years, but neither age nor BMI had an effect on dancing ability (both $F_{1,34} < 2.25$, $P > 0.15$).

To test for sex differences between evaluators in the strength of preference for symmetrical individuals, a variable called 'relative preference for symmetrical dancers' was constructed by subtracting dance evaluations given to asymmetrical dancers from the dance evaluations given to symmetrical dancers. Higher scores indicated stronger preferences for symmetrical dancers relative to asymmetrical dancers. Female evaluators had a stronger relative preference (Fig. 1b) for symmetrical male dancers (20.43 ± 13.54) than male evaluators (14.90 ± 17.55) ($t_{154} = 2.21$, $P = 0.029$), while there was no sex difference in dance ratings of symmetrical females ($t_{154} = 1.50$, $P = 0.137$). However, male evaluators did give higher ratings to the dances of females (43.75 ± 17.67) than did female evaluators (37.87 ± 15.08) ($t_{154} = 2.19$, $P = 0.03$).

Does fluctuating asymmetry of the evaluator shift preferences away from symmetrical individuals? A multiple regression analysis revealed the predicted association in men (Fig. 2). That is, FA in male evaluators was negatively associated with relative preference for dances performed by symmetrical females. Specifically, changes in male composite relative FA accounted for 11% of the differences in

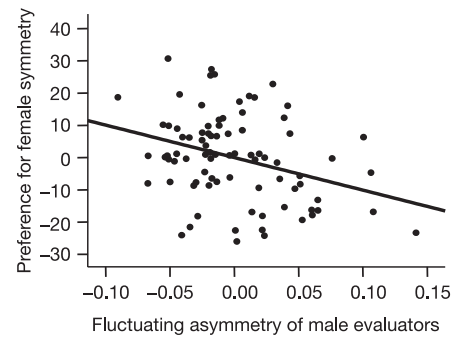


Figure 2 | Male evaluator FA and preferences for female symmetry. A partial regression plot (age and body mass index controlled) showing a negative relationship between male evaluators' composite relative fluctuating asymmetry and preference for symmetrical female dancers. Both variables in the partial regression plot are residuals.

the relative preference for the dances of symmetrical females when age and BMI were included as covariates (partial $R^2 = 0.11$, $P = 0.02$). There was no significant association between female evaluator FA and preferences for symmetrical males' dances, $R^2 = 0.02$, $P = 0.32$.

We do not know what mediates the associations reported—perhaps asymmetry itself or a covarying characteristic such as neuromuscular coordination or health, including freedom from parasites^{2,29}. Attractive dances may be more difficult to perform, more rhythmic, more energetic, more energy efficient or any combination of these factors. As motion-capture technology stores each dance in a mathematical form, we hope to discover more precisely which patterns of dance movement are associated with both quality of dance and FA of the dancer. Does dance ability correlate with reproductive success? We plan to address this question with long-term data from the same population.

METHODS

Fluctuating asymmetry. Morphometric measurements were collected for each dancer (wrists, ankles, elbows, third digit, fourth digit, fifth digit, ears, feet, knees) with vernier calipers to 0.01 mm accuracy (see the methodology in ref. 9) in 1996 and 2002 in Southfield, Jamaica (refer to the Supplementary Background and Analyses for details). To establish repeatability levels and reduce measurement error, each trait was measured twice and averaged⁹. Bilateral trait measurements were found to be reliable indicators of between-subject differences (as opposed to measurement error), and reflect true FA rather than biologically significant directional asymmetry or antisymmetry (that is, the signed trait asymmetries did not show significant platykurtosis, an indicator of antisymmetry rather than true FA⁹). Composite relative FA was calculated by subtracting the average length of the right side of the trait from the left (L minus R) and correcting for trait size³⁰, and then summing the absolute values across all traits.

Motion capture. A large sample ($n = 183$) were filmed with motion-capture cameras under constant conditions: one minute, same music, and danced within a 4-m² space. An optical motion-capture system from Vicon was chosen owing to the freedom provided to the subject when using lightweight markers. Optical motion-capture systems have been extensively applied in biomechanics for gait analysis and motion research. The system used eight cameras (120-frames-per-second capture rate) in order to capture large amounts of dance movement while removing potential visual confounds (including body size, attractiveness, clothing and identity). These cameras (Supplementary Fig. 1) tracked movements using 41 infrared reflectors (Supplementary Fig. 2). Subjects wore relatively tight clothing to facilitate accurate information from the markers. Reflector trajectories were used to robustly reconstruct joint angles of the three-dimensional (3D) skeletal animation, and transformed this information into a virtual 3D animation. To post-process captured raw data, Vicon IQ 1.5 software was used. The animations were displayed with visualization software developed at the University of Washington. This software used accepted techniques for generating accurate animations from marker positions.

Stimuli presentation. Forty dance animations were selected (20 symmetrical individuals (10 female, 10 male), mean age = 17.89 ± 1.84 , and 20 asymmetrical individuals (10 female, 10 male), mean age = 17.40 ± 1.79). The

criterion for selection was that a symmetrical individual must have been in the bottom third percentile of fluctuating asymmetry in both 1996 and the 2002 re-measure (the reverse was used for the asymmetrical category). There were no statistically significant sex differences in mean or variances in FA between the sexes within a given category (Supplementary Background and Analyses). Despite the subtle differences in FA between asymmetrical and symmetrical dancers, the means were statistically different, all t values were >11.00 , all P values were <0.001 . The 40 dance animations with audio were randomly presented on a 2-m white background sequentially using an InFocus DLP Projector to a sample of 155 young adults (mean age = 18.12 ± 1.75 ; 87 males, 68 females), also part of the Jamaican Symmetry Project⁸, for evaluation on a 90-mm scale with the left side labelled "Bad Dancer" and the right side labelled "Good Dancer". Each subject was asked to place a vertical mark on this scale. Marks were measured to the nearest millimetre by a research assistant blind with respect to condition. Each subject was also asked to identify the sex of the dancer in a forced-choice question. To restrict the study to evaluations of others, we removed 31 self-evaluations (15 male and 16 female) where individuals rated their own dances. The virtual 3D dance animations were presented in an approximate front-facing angle by an experimenter blind with respect to condition and hypotheses. Presentation order was not related to dance evaluations (Pearson $r = 0.04$, $P = 0.82$) or dancer FA (point biserial $r = -0.14$, $P = 0.40$).

Statistical analyses. A 2 (Sex of Dancer) $\times$ 2 (Symmetry of Dancer) between-subjects ANCOVA (analysis of covariance), with dance ability as the dependent variable, tested the primary hypothesis that symmetrical individuals are better dancers (data analysed by SPSS 12.0 from SPSS Inc.). BMI and age were included as covariates. Dance ability variances were not significantly different, although a marginal difference was observed whereby males show greater variability in dance ability (180.19) than do females (113.03) (Levene test $F = 3.29$, $P = 0.08$). To investigate individual differences among evaluators in the strength of symmetry preference, t -tests were used. Results remain significant when using a nonparametric Mann–Whitney test for heterogeneous variances. Multiple regression analysed whether or not fluctuating asymmetry of evaluators was negatively associated with relative preference for symmetrical dancers. Age and BMI were entered as covariates to minimize the influence of these potential confounds. Neither age nor BMI of dancers nor evaluators were associated with dance ability or the evaluators' preferences. Finally, the effect of bodily symmetry on dance ability is independent of facial attractiveness and self-esteem of the dancer (Supplementary Background and Analyses).

Received 23 September; accepted 20 October 2005.

1. Darwin, C. R. *The Descent of Man and Selection in Relation to Sex* (D. Appleton and Company, New York, 1871/1896).
2. Brønseth, T. & Folstad, I. The effect of parasites on courtship dance in threespine sticklebacks: More than meets the eye? *Can. J. Zool.* **75**, 589–594 (1997).
3. Singer, F. et al. Analysis of courtship success in the funnel-web spider *Agelenopsis aperta*. *Behaviour* **137**, 93–117 (2000).
4. Rintamaki, P. T., Alatalo, R. V., Hoglund, J. & Lundberg, A. Fluctuating asymmetry and copulation success in lekking black grouse. *Anim. Behav.* **54**, 265–269 (1997).
5. Hanna, J. L. *Dance, Sex, and Gender: Signs of Identity, Dominance, Defiance, and Desire* (Univ. of Chicago Press, Chicago, 1988).
6. Grammer, K., Keki, V., Striabel, B., Atzmueller, M. & Fink, B. in *Evolutionary Aesthetics* (eds Volland, E. & Grammer, K.) 295–324 (Springer, Heidelberg, 2003).
7. Polak, M. (ed.) *Developmental Instability: Causes and Consequences* (Oxford Univ. Press, New York, 2003).
8. Møller, A. P. & Swaddle, J. P. *Developmental Stability and Evolution* (Oxford Univ. Press, Oxford, 1997).
9. Trivers, R., Manning, J. T., Thornhill, R., Singh, D. & McGuire, D. Jamaican Symmetry Project: Long-term study of fluctuating asymmetry in rural Jamaican children. *Hum. Biol.* **71**, 417–430 (1999).
10. Van Valen, L. A study of fluctuating asymmetry. *Evolution* **16**, 125–142 (1962).
11. Waddington, C. H. *The Strategy of the Genes* (Macmillan, New York, 1957).
12. Martin, J. & Lopez, P. Hindlimb asymmetry reduces escape performance in the lizard *Psammotromus algerius*. *Physiol. Biochem. Zool.* **74**, 619–624 (2001).
13. Møller, A. P., Sanotra, G. S. & Vestergaard, K. S. Developmental instability and light regime in chickens (*Gallus gallus*). *Appl. Anim. Behav. Sci.* **62**, 57–71 (1999).
14. Swaddle, J. P. Within-individual changes in developmental stability affect flight performance. *Behav. Ecol.* **8**, 601–604 (1997).
15. Garland, T. Jr & Freeman, P. W. Selective breeding for high endurance running increases hindlimb symmetry. *Evolution* **59**, 1851–1854 (2005).
16. Gummer, D. L. & Brigham, R. M. Does fluctuating asymmetry reflect the importance of traits in little brown bats (*Myotis lucifugus*)? *Can. J. Zool.* **73**, 990–992 (1995).
17. Moodie, G. E. E. Meristic variation, asymmetry, and aspects of the habitat of *Culea inconstans* (Kirtland), the brook stickleback, in Manitoba. *Can. J. Zool.* **55**, 398–404 (1977).
18. Manning, J. T. & Ockenden, L. Fluctuating asymmetry in racehorses. *Nature* **370**, 185–186 (1994).
19. Manning, J. T. & Pickup, L. J. Symmetry and performance in middle distance runners. *Int. J. Sports Med.* **19**, 205–209 (1998).
20. Thornhill, R. & Gangestad, S. The scent of symmetry: A human sex pheromone that signals fitness? *Evol. Hum. Behav.* **20**, 175–201 (1999).
21. Hughes, S. M., Harrison, M. A. & Gallup, G. G. The sound of symmetry: voice as a marker of developmental instability. *Evol. Hum. Behav.* **23**, 173–180 (2002).
22. Gangestad, S. W., Thornhill, R. & Yeo, R. A. Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethol. Sociobiol.* **15**, 73–85 (1994).
23. Palmer, A. R. Detecting publication bias in meta-analyses: A case study of fluctuating asymmetry and sexual selection. *Am. Nat.* **154**, 220–233 (1999).
24. Josefsson, T., Nordh, E. & Eriksson, P. O. A flexible high-precision video system for digital recording of motor acts through lightweight reflex markers. *Comput. Methods Programs Biomed.* **49**, 119–129 (1996).
25. Nagata, N., Okumoto, K., Iwai, D., Toro, F. & Inokuchi, S. Analysis and synthesis of Latin dance using motion capture data. *Lect. Notes Comput. Sci.* **3333**, 39–44 (2004).
26. Kellner, J. & Alford, R. A. The ontogeny of fluctuating asymmetry. *Am. Nat.* **161**, 931–947 (2003).
27. Trivers, R. in *Sexual Selection and the Descent of Man 1871–1971* (ed. Campbell, B.) 136–179 (Aldine Publishing, Chicago, 1972).
28. Little, A. C., Burt, D. M., Penton-Voak, I. S. & Perrett, D. I. Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proc. R. Soc. Lond. B* **268**, 39–44 (2001).
29. Hamilton, W. D. & Zuk, M. Heritable true fitness and bright birds: a role for parasites? *Science* **218**, 384–387 (1982).
30. Palmer, A. R. & Strobeck, C. Fluctuating asymmetry: Measurement, analysis and pattern. *Annu. Rev. Ecol. Syst.* **17**, 391–421 (1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Jamaican Ministry of Education and Culture for permission to conduct research. We are grateful for assistance from teachers, principals, parents, students, M. Cuff, B. Dunham, N. Sutherland and D. Zaatar. Financial support was provided by Rutgers University, the University of Washington Animation Labs, the Ann and Gordon Getty Foundation, the Rutgers Center for Human Evolutionary Studies, the Biosocial Research Foundation, and NSF grants awarded to L.C., Z.P. and R.T. W.M.B. was supported by an NSERC (Canada) postdoctoral fellowship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.M.B. (wmbrown@rci.rutgers.edu).

Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*

William C. Nierman¹, Arnab Pain², Michael J. Anderson³, Jennifer R. Wortman¹, H. Stanley Kim¹, Javier Arroyo⁴, Matthew Berriman², Keietsu Abe⁵, David B. Archer⁶, Clara Bermejo⁴, Joan Bennett⁷, Paul Bowyer³, Dan Chen¹, Matthew Collins², Richard Coulsen⁸, Robert Davies², Paul S. Dyer⁶, Mark Farman⁹, Nadia Fedorova¹, Natalie Fedorova¹, Tamara V. Feldblyum¹, Reinhard Fischer¹⁰, Nigel Fosker², Audrey Fraser², Jose L. García¹¹, Maria J. García¹², Arlette Goble², Gustavo H. Goldman¹³, Katsuya Gomi⁵, Sam Griffith-Jones², Ryan Gwilliam², Brian Haas¹, Hubertus Haas¹⁴, David Harris², H. Horiuchi¹⁵, Jiaqi Huang¹, Sean Humphray², Javier Jiménez¹², Nancy Keller¹⁵, Hoda Khouri¹, Katsuhiko Kitamoto¹⁶, Tetsuo Kobayashi¹⁷, Sven Konzack¹⁰, Resham Kulkarni¹, Toshitaka Kumagai¹⁸, Anne Lafton¹⁹, Jean-Paul Latgé²⁰, Weixi Li⁹, Angela Lord², Charles Lu¹, William H. Majoros¹, Gregory S. May²¹, Bruce L. Miller²², Yasmin Mohamoud¹, Maria Molina⁴, Michel Monod²³, Isabelle Mouyna²⁰, Stephanie Mulligan¹, Lee Murphy², Susan O'Neil², Ian Paulsen¹, Miguel A. Peñalva¹¹, Mihaela Pertea¹, Claire Price², Bethan L. Pritchard³, Michael A. Quail², Ester Rabbinoiwitsch², Neil Rawlins², Marie-Adele Rajandream², Utz Reichard²⁴, Hubert Renauld², Geoffrey D. Robson³, Santiago Rodriguez de Córdoba¹¹, Jose M. Rodríguez-Peña⁴, Catherine M. Ronning¹, Simon Rutter², Steven L. Salzberg¹, Miguel Sanchez¹², Juan C. Sánchez-Ferrero¹¹, David Saunders², Kathy Seeger², Rob Squares², Steven Squares², Michio Takeuchi²⁵, Fredj Tekaia²⁰, Geoffrey Turner²⁶, Carlos R. Vazquez de Aldana¹², Janice Weidman¹, Owen White¹, John Woodward², Jae-Hyuk Yu²⁷, Claire Fraser¹, James E. Galagan²⁸, Kiyoshi Asai¹⁸, Masayuki Machida²⁹, Neil Hall^{2†}, Bart Barrell² & David W. Denning³

Aspergillus fumigatus is exceptional among microorganisms in being both a primary and opportunistic pathogen as well as a major allergen^{1–3}. Its conidia production is prolific, and so human respiratory tract exposure is almost constant⁴. *A. fumigatus* is isolated from human habitats⁵ and vegetable compost heaps^{6,7}. In immunocompromised individuals, the incidence of invasive infection can be as high as 50% and the mortality rate is often about 50% (ref. 2). The interaction of *A. fumigatus* and other airborne fungi with the immune system is increasingly linked to severe asthma and sinusitis⁸. Although the burden of invasive disease caused by *A. fumigatus* is substantial, the basic biology of the organism is mostly obscure. Here we show the complete 29.4-megabase genome sequence of the clinical isolate Af293, which consists of eight chromosomes containing 9,926 predicted genes. Microarray analysis revealed temperature-dependent expression

of distinct sets of genes, as well as 700 *A. fumigatus* genes not present or significantly diverged in the closely related sexual species *Neosartorya fischeri*, many of which may have roles in the pathogenicity phenotype. The Af293 genome sequence provides an unparalleled resource for the future understanding of this remarkable fungus.

The genome of *A. fumigatus* Af293 was sequenced by the whole-genome random sequencing method⁹ augmented by optical mapping¹⁰. Genome closure and quality standard attainment was accomplished by directed sequencing and manual editing. (See Table 1 and Supplementary Fig. S1 for genome features.) Sequenced chromosomal arms extend from putative centromeres to the telomere and end in 7–21 tandem repeats of the sequence TTAGGG. The copy number of the mitochondrial genome relative to the nuclear genome is estimated to be 12 based on the redundancy in the assembled

¹The Institute for Genomic Research, Rockville, Maryland 20850, and The George Washington University School of Medicine, Department of Biochemistry and Molecular Biology, 2300 Eye Street NW, Washington DC 20037, USA. ²The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. ³School of Medicine and Faculty of Life Sciences, The University of Manchester, Stopford Building, Manchester M13 9PT, UK. ⁴Departamento Microbiología II. Universidad Complutense de Madrid 28040, Spain. ⁵Tohoku University, 1-1 Tsutsumidori-Amamiyamachi Aoba-ku, Sendai 981-8555, Japan. ⁶School of Biology, University of Nottingham, University Park, Nottingham NG7 2RD, UK. ⁷Department of Cell and Molecular Biology, Tulane University, New Orleans, Louisiana 70118, USA. ⁸European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK. ⁹Department of Plant Pathology, University of Kentucky, Lexington, Kentucky 40546-0312, USA. ¹⁰Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany. ¹¹Centro de Investigaciones Biológicas, CSIC, Madrid 28040, Spain. ¹²Departamento Microbiología y Genética, Universidad de Salamanca, 37007 Salamanca, Spain. ¹³Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Brazil. ¹⁴Department of Molecular Biology, Innsbruck Medical University, A-6020 Innsbruck, Austria. ¹⁵Department of Plant Pathology, University of Wisconsin at Madison, Madison, Wisconsin 53706, USA. ¹⁶Department of Biotechnology, University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan. ¹⁷Department of Biological Mechanisms and Functions, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan. ¹⁸Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan. ¹⁹Unité Postulante Biologie et Pathogénicité Fongiques, INRA USC 2019, Institut Pasteur, Paris 75015, France. ²⁰Unité des Aspergillus, Institut Pasteur, Paris 75015, France. ²¹Division of Pathology and Laboratory Medicine, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA. ²²Departments of Microbiology, Molecular Biology and Biochemistry, Center for Reproductive Biology, University of Idaho, Moscow, Idaho 83844, USA. ²³Department of Dermatology, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland. ²⁴Department of Bacteriology, Georg-August-University, D-37077 Göttingen, Germany. ²⁵Tokyo University of Agriculture and Technology, Saiwai-chou 3-5-8, Fuchu, Tokyo 183-0054, Japan. ²⁶Department of Molecular Biology and Biotechnology, The University of Sheffield, Sheffield S10 2TN, UK. ²⁷Department of Food Microbiology and Toxicology, The University of Wisconsin, Madison, Wisconsin 53706, USA. ²⁸Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139, USA. ²⁹Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8566, Japan. [†]Present address: The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA.

sequence. The protein-coding genes and other genome features were identified by an automated annotation pipeline coupled with manual review.

Several candidate pathogenicity genes have been previously identified by assaying mutants in cultured macrophages or in animal models of invasive aspergillosis. These genes encode proteins involved in central metabolic pathways, signalling, cell wall biosynthesis, pigment biosynthesis and regulation of secondary metabolite production (Supplementary Table S1). This scope of functions suggests that the genomic infrastructure for pathogenicity is complex and integrated with a range of metabolic capabilities. Thus, any computationally based analysis of the genome sequence would not be directly able to identify functions critical for pathogenicity.

A. fumigatus thermotolerance is a trait critical to its ability to thrive in mammalian and avian infections and in the even-higher temperature ranges characteristic of composts (that is, up to 70 °C). To investigate the metabolic adaptation of this fungus to higher temperatures, gene expression was examined throughout a time course upon shift of growth temperatures from 30 °C (representing environments of tropical soil) to 37 °C and 48 °C (representing temperatures in the human body and compost, respectively). Gene expression patterns revealed that comparable numbers of genes were differentially expressed at each temperature, many of them with similar patterns (Fig. 1a). We identified 323 genes (clusters 1 and 2) that showed a higher expression level at 48 °C than at 37 °C, and 135 genes (cluster 3) that were expressed at a higher level at 37 °C than at 48 °C (Fig. 1b, see also Supplementary Table S2). Many of these 323 genes, especially those in cluster 1, which is enriched with heat shock-responsive genes, may have a role in thermotolerance of *A. fumigatus*. These include only 11 (four in cluster 1 and seven in cluster 2) of the 551 homologues of the *Saccharomyces cerevisiae* general stress-response genes, which were shown to be differentially expressed under all stress conditions tested¹¹. Cluster 3 also includes a small number of such genes (five), and three of them have the opposite expression patterns from yeast (Supplementary Table S2). These data indicate that high temperature responses in *A. fumigatus* differ from

the general stress response in yeast. Except for catalase B, no known genes implicated in pathogenicity showed higher expression at 37 °C than at 48 °C, suggesting that host temperature alone (37 °C) is insufficient to turn on many virulence-related genes.

More allergens (defined by IgE binding) have been characterized from *A. fumigatus* than from all other fungal species combined ($n = 58$)¹². We identified nine additional predicted allergens in the genome based on similarity with other fungal allergens (Supplementary Table S3), including secreted proteases, glucanases and cellulases. Only *A. fumigatus* encodes the major allergen ribotoxin (Asp f1), which cleaves a single phosphodiester bond of the 28S ribosomal RNA of eukaryotic ribosomes. None of the nine allergens is a spore surface protein, despite a hydrophobin in *Cladosporium herbarum* being allergenic¹³. The allergen Asp f16 has immunoprotective properties¹⁴.

Identification of essential genes may reveal potential targets for drug development. Putative essential genes in the *A. fumigatus* genome were identified by BLASTp search against 131 single-member KOGs (eukaryotic orthologous groups) representing a conserved core of largely essential eukaryotic genes compiled by analysis of seven diverse, completely sequenced eukaryotic genomes¹⁵ (Supplementary Table S4). Only one of the 131 KOGs, KOG3214/DUF701, containing putative Zn ribbon RNA binding proteins, was not found in *A. fumigatus* or other aspergilli, suggesting a lineage-specific gene loss.

A. fumigatus virulence may be augmented by its numerous secondary metabolites, including fumagillin, gliotoxin, fumitremorgin, verruculogen, fumigaclavine, helvolic acid and sphingofungins⁴. Genes controlling fungal secondary metabolites are generally organized in clusters, many of which are species-specific. The *A. fumigatus* genome contains 26 such clusters with polyketide synthase, non-ribosomal peptide synthase and/or dimethylallyl tryptophan synthase genes. Only 13 of the 26 clusters have orthologues in *A. oryzae* and/or *A. nidulans*, and ten of these orthologous clusters are missing many or most of the genes present in the *A. fumigatus* clusters (see 'Selfish Cluster Hypothesis' in Supplementary Information). The unique clusters of *A. fumigatus* are dispersed in the genome with a bias towards telomeric locations. Many of these clusters contain regulatory genes, genes associated with resistance such as transporters involved in efflux¹⁶, and genes with no obvious role in production of the metabolite (Supplementary Table S5). Fifteen of the clusters contain 22 transcriptional regulators, which are probably specific to their cluster because they do not have strong similarity to other proteins in the databases. In contrast to these regulators within the clusters, other global regulators of secondary metabolite synthesis are dispersed in the *A. fumigatus* genome. The genome also contains one copy of *laeA* encoding a global regulator of *Aspergillus* secondary metabolites¹⁷.

Table 2 summarizes the numbers of different classes of secondary metabolite genes for *A. fumigatus*, *A. nidulans* (ref. 18) and *A. oryzae* (ref. 19). (See 'Secondary Metabolites' in Supplementary Information for further discussion.)

Stimulation of the programmed cell death pathway, as reported for *A. fumigatus* and *A. nidulans* during stationary phase and oxidative death²⁰, presents an opportunity for antifungal drug development.

Table 1 | Properties of the *Aspergillus fumigatus* Af293 genome

Genome	Value
Nuclear genome	
General information	
Size (Mb)	29.4
G+C content (%)	49.9
Gene number	9,926
Mean gene length (bp)	1,431
Per cent coding	50.1
Per cent genes with introns	77.0
Genes of unknown function	3,288
Exons	
Mean number per gene	2.8
Mean length (bp)	516
G+C content (%)	54.0
Introns	
Mean number per gene	1.8
Mean length (bp)	112
G+C content (%)	46.3
Intergenic regions	
Mean length (bp)	1,226
G+C content (%)	46.0
RNA	
tRNA number	179
5S rRNA number	33
Mitochondrial genome	
Size (bp)	31,892
G+C content (%)	25.4
Gene number	16
Mean gene length (bp)	1,189
Per cent coding	44.1
Per cent genes with introns	6.2
tRNA number	33

Table 2 | Secondary metabolite gene types in *A. fumigatus*, *A. nidulans* and *A. oryzae*

Gene type	<i>A. fumigatus</i>	<i>A. nidulans</i>	<i>A. oryzae</i>
Polyketide synthase	14	27	30
Non-ribosomal peptide synthase	14†	14†	18
Fatty acid synthase*	1	6	5
Sesquiterpene cyclase	n.d.	1	1
Dimethylallyl tryptophan synthase	7	2	2

* Includes one primary metabolism fatty acid synthase in each species.

† Includes one hybrid polyketide/peptide synthase.

n.d., not detected.

As with other fungi, *A. fumigatus* lacks homologues of the metazoan upstream apoptotic machinery, whereas the downstream effectors and regulators, both caspase-dependent and caspase-independent, seem to be shared (Supplementary Table S6). *A. fumigatus* possesses a

homologue of the key participant of caspase-independent apoptosis in mammals, PARP, which is absent in *S. cerevisiae*. PARP activity was demonstrated previously in *A. nidulans* during sporulation-induced apoptosis²¹. The presence of these proteins in *Aspergillus* is indicative

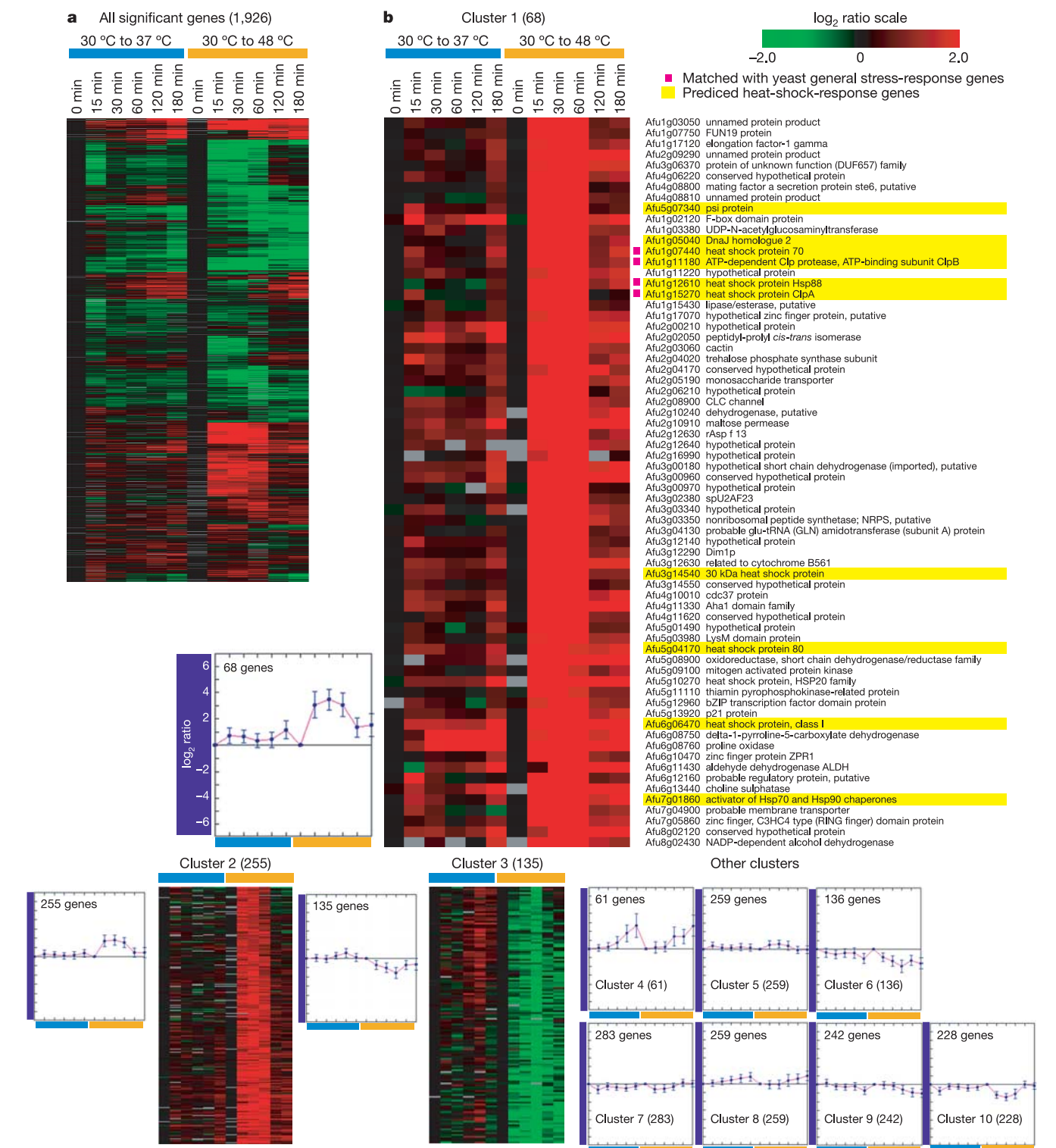


Figure 1 | Gene expression profiles for the temperature shift-responsive genes. The colour bar indicates the range of the expression ratios in the heat map-type figures. **a**, Genes with significantly differentiated expression are shown (see Methods). **b**, The same set of genes grouped into ten clusters. Three clusters of interest (that is, clusters 1 and 2 with genes expressed at a higher level at 48 °C than at 37 °C, and cluster 3 with the opposite pattern)

are shown with centroid graphs and heat map-type figures. Cluster 1 enriches heat shock genes that are highlighted in yellow. Homologues to the yeast general stress-response genes are indicated with pink boxes. Vertical bars on graphs represent the data range, and the points in the middle the average.

of the recently identified PARP-dependent programmed cell death pathway and makes these filamentous fungi attractive models in which to study the mechanism and origin of programmed cell death.

As the hyphal cell wall is essential for *A. fumigatus* to penetrate solid nutrient substrates and to resist host cell defence reactions, comprehension of cell wall biosynthesis pathways are important. The *A. fumigatus* cell wall is composed of a fibrillar branched β 1,3-glucan core bound to chitin, galactomannan and β 1,3-1,4-glucan, embedded in an amorphous cement composed of α 1,3-glucan, galactomannan and polygalactosamine²². β 1,6-glucan and peptidomannan, both present in yeast cell walls, are missing in *A. fumigatus*. The types and numbers of *A. fumigatus* Af293 cell wall-related proteins as compared to other eukaryotes are provided in Supplementary Table S7. Specificity of the polymer organization of the *A. fumigatus* cell wall is reflected at the genomic level in the specificity of the cell wall biosynthetic gene inventory.

In *S. cerevisiae*, certain proteins initially anchored by a glycosyl phosphatidylinositol (GPI) moiety to the plasma membrane, and subsequently cross-linked to β 1,3-glucans through β 1,6-glucans, are thought to be major participants in yeast cell wall organization²³. Among 82 putative GPI-anchored proteins identified in *A. fumigatus*, no homologues of these yeast GPI-anchored proteins were found (Supplementary Fig. S2). *A. fumigatus* also lacked homologues of the yeast PIR proteins that are putatively bound to the β 1,3-glucans through an alkali-labile bond. It has been hypothesized in yeast that the linkage of proteins to cell wall polysaccharides is important in establishing the three-dimensional polysaccharide network that constitutes the skeleton of all fungal cell walls. On the basis of the

comparative analysis reported here, it is more likely that binding to polysaccharides in yeasts is merely a way for certain proteins to remain at the surface of the cell wall to fulfil their biological functions in adhesion and flocculation—events absent in mould biology—and in mating. Hydrophobins, proteins not found in *S. cerevisiae*, are the only cell wall-linked GPI proteins detected in the *A. fumigatus* genome sequence. Hydrophobins have a major role in mould biology, because they are required for attachment to hydrophobic surfaces, formation of aerial structures, air dispersion and survival of conidia.

More than 500 putative *A. fumigatus*-specific genes having no detectable *A. nidulans* or *A. oryzae* homologues were found, mostly annotated as hypothetical proteins. *A. fumigatus*-specific proteins that have functional annotations other than hypothetical are listed in Supplementary Table S8. Most of these seem to have unusual phyletic patterns and are clustered in synteny break locations relative to *A. oryzae* and *A. nidulans* (Fig. 2 in ref. 18). About one-third of the *A. fumigatus*-specific proteins showed significant similarity to other fungal gene products. Furthermore, many seem to be involved in secondary metabolite biosynthesis, such as the developmentally regulated cluster involved in conidial pigment biosynthesis in *A. fumigatus*.

Several of the *A. fumigatus*-specific genes apparently have only bacterial or archaeal homologues, and may confer selective advantage in adapting to environments as diverse as human bodies, compost piles and arsenic-contaminated soil. The most striking finding involves two *A. fumigatus*-specific proteins that show high sequence similarity with the pI258 ArsC superfamily of arsenate reductases,

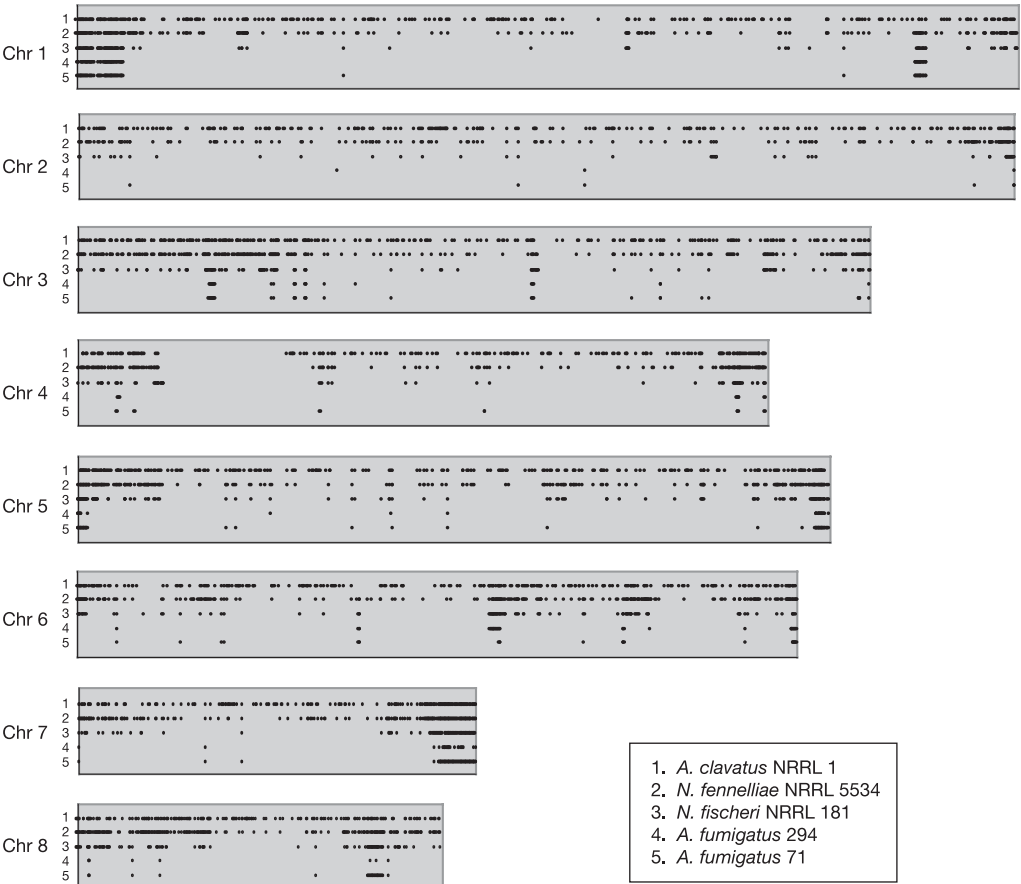


Figure 2 | Spatial distribution of *A. fumigatus* Af293 genes not present or diverged as compared with various (unsequenced) strains. On the basis of the microarray CGH data, *A. fumigatus* Af293 genes (reference) with log₂ ratios equal to or greater than 2 as compared to signals from the query

strains are scored as absent or diverged in the query strains. The five query strains are denoted with the numbers 1–5. The locations of *A. fumigatus* genes for which the orthologues are diverged or missing in the query strain are arranged in the order that they appear along the chromosome.

responsible for detoxification of arsenate by reduction to arsenite in bacteria²⁴. These two proteins are unrelated to Acr2p of *S. cerevisiae* and are the first instances of the p1258 ArsC-type arsenate reductase in eukaryotes. The corresponding *A. fumigatus* genes are in a duplicated cluster on chromosomes 1 and 5, along with genes encoding an arsenite exporter, an arsenic resistance protein and an arsenic methyltransferase (Supplementary Table S9). It is of particular note that the cluster members seem to have different phylogenetic patterns. Although all of the significant BLASTp hits for the arsenate reductase and arsenic resistance protein are actinobacterial and proteobacterial proteins, the arsenite exporter appears to be closely related to yeast Acr3p and the methyltransferase has significant similarity to *Neurospora crassa* and archaeal proteins as well as mammalian S-adenosyl-L-methionine:AsIII methyltransferase²⁵. The selective benefits of the assembly and retention of this cluster may involve the co-regulation of these arsenic resistance genes²⁶. Elsewhere in the *A. fumigatus* genome, genes for an arsenite efflux pump and an arsenite translocating ATPase as well as additional copies of the arsenate exporter and arsenic resistance genes have been identified (Supplementary Table S9). This gene complement supports the classification of *A. fumigatus* among the once notorious 'arsenic fungi', organisms that produce the volatile trimethylarsine when grown in arsenate-contaminated environments²⁷.

The genome sequence of *A. fumigatus* revealed several genes associated with mating processes and sexual development. This topic is discussed further in an accompanying paper¹⁸.

Azoles and allylamines block two sequential steps in the 20-step cascade of ergosterol synthesis. Comparative analysis of the ergosterol synthesis pathway genes revealed variable copy numbers of several genes, including *ERG3* and *ERG11* (Supplementary Table S10). Duplicated genes in the *Aspergillus* ERG pathway may reflect an adaptation strategy modulating the composition and fluidity of the cell membrane.

The comparative analysis of the *A. fumigatus* genome has made good use of the sequences of the *A. nidulans* and *A. oryzae* genomes to study gene and genome evolution among these species (see the accompanying paper¹⁸). However, within the genus *Aspergillus*, *A. nidulans* and *A. oryzae* are only distantly related to *A. fumigatus*. To explore the association between gene content and phenotype (that is, pathogenicity and related subphenotypes) the much closer taxonomic relationship of *Neosartorya fischeri* and *Neosartorya fennelliae* to *A. fumigatus* provides a more powerful comparative set. *N. fennelliae* is not known to be pathogenic to humans and possesses a sexual cycle. Another closely related species is *Aspergillus clavatus*, a mycotoxin producer that has been implicated in neurotoxicosis in beef cattle as well as respiratory disease in maltworkers²⁸. We have used genomic DNA from *N. fischeri*, *N. fennelliae* and *A. clavatus* as well as from two additional strains of *A. fumigatus*, Af294 and Af71, to perform comparative genomic hybridization (CGH) with our Af293 polymerase chain reaction (PCR) amplicon coding sequence (CDS) microarray. The analysis revealed 2,557 total *A. fumigatus* Af293 genes to be absent or diverged in the analysed species. Of these, 1,382 are assigned gene names, including 70 coding for enzymes involved in transcriptional regulation, at least 22 in production of secondary metabolites, and 6 encoding proteins for drug resistance transporters. Both of the *arsC* genes were missing or diverged in most of the analysed strains, including *A. fumigatus* strains Af294 and Af71. Figure 2 shows the chromosomal locations of the missing or diverged genes, demonstrating a bias towards subtelomeric locations consistent with the higher density of synteny breaks observed in subtelomeric locations between *A. fumigatus*, *A. nidulans* and *A. oryzae* (Fig. 2 of ref. 18) and suggesting greater genome instability in these regions. The most relevant CGH analysis for phenotypic comparisons, that with *N. fischeri*, revealed 700 genes to be absent or diverged relative to *A. fumigatus* Af293 (Supplementary Table S11). These include at least 13 genes coding for enzymes involved in the production of secondary metabolites, 28 coding

for transcriptional regulators and protein kinases, 21 coding for transporters, 199 coding for metabolic and other proteins, and 400 coding for hypothetical proteins. This number of genes is a manageable set to begin the effort of correlating phenotypic differences between these species to gene content.

METHODS

Strain isolates. Af293 was isolated from a patient who ultimately died from invasive aspergillosis²⁹. Af71 (NCPF 7098) and Af294 (NCPF 7102) are also clinical isolates. The type strains of *N. fischeri* (NRRL 181), *N. fennelliae* (NRRL 5534) and *A. clavatus* (NRRL 1) were used for CGH.

Sequencing and assembly. The genome of *A. fumigatus* Af293 was sequenced and assembled using the random shotgun method. Closure (finishing) was accomplished by directed sequencing and manual editing of the genome sequence⁹. Sequencing and assembly statistics are provided in Supplementary Methods.

Coding sequence prediction and gene identification. The assembled genomic sequence was processed through the TIGR annotation pipeline, a collection of software known as Eukaryotic Genome Control (EGC) that serves as the central data management system. This pipeline is described in detail in Supplementary Methods.

Microarray methods. The DNA amplicon microarray for *A. fumigatus* Af293 was constructed by designing primers for 9,516 genes (96%) then amplifying these target gene regions from genomic DNA (see Supplementary Methods). The resulting PCR products were purified and spotted in triplicate at high density on Corning UltraGAPS aminosilane-coated microscope slides using a robotic spotter built by Intelligent Automatic Systems and cross-linked by ultraviolet illumination.

For CGH analyses, genomic DNA was prepared from each isolate using the DNeasy Tissue kit (Qiagen). Purified genomic DNA was labelled and hybridized as described³⁰. For temperature-shift experiments, conidia ($5 \times 10^6 \text{ ml}^{-1}$) from Af293 were incubated in Complete medium for germination ($\sim 17 \text{ h}$) at 30°C . Cultures were then transferred to a water bath of 37°C or 48°C for continued growth. Total RNA samples before (that is, 0 min) and after (that is, 15, 39, 60, 120 and 180 min) two temperature shifts (that is, 30 to 37°C and 30 to 48°C) were used to profile gene expression. A biological replication of the cell growths and samplings was conducted. Labelling reactions with RNA and hybridizations were conducted as described in the TIGR standard operating procedures found at <http://atarrays.tigr.org>. The sample from 0 min in each temperature-shift set served as a reference in all hybridizations with samples from later time points within the set. All of the hybridizations with the two biological replicates were repeated in dye-swap sets.

Hybridized slides were scanned and analysed to obtain relative transcript levels (see Supplementary Methods). Normalized data were averaged over replications, and differentially expressed genes at the 95% confidence level were determined using intensity-dependent Z-scores (with $Z = 1.96$). The resulting data were organized and visualized using euclidean distance and hierarchical clustering with average linkage clustering method to view the whole data set (Fig. 1a) and k-means to group the genes in ten clusters (Fig. 1b) with TIGR MEV (<http://www.tigr.org/software>).

Received 12 May; accepted 12 October 2005.

- Casadevall, A. & Pirofski, L. A. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infect. Immun.* **67**, 3703–3713 (1999).
- Denning, D. W. Invasive aspergillosis. *Clin. Infect. Dis.* **26**, 781–803, 804–805 (1998).
- Greenberger, P. A. Allergic bronchopulmonary aspergillosis. *J. Allergy Clin. Immunol.* **110**, 685–692 (2002).
- Latge, J. P. *Aspergillus fumigatus* and aspergillosis. *Clin. Microbiol. Rev.* **12**, 310–350 (1999).
- Hirsch, T. et al. House-dust-mite allergen concentrations (Der f 1) and mold spores in apartment bedrooms before and after installation of insulated windows and central heating systems. *Allergy* **55**, 79–83 (2000).
- Beffa, T. et al. Mycological control and surveillance of biological waste and compost. *Med. Mycol.* **36** (suppl. 1), 137–145 (1998).
- Ryckebøer, J., Mergaert, J., Coosemans, J., Depriens, K. & Swings, J. Microbiological aspects of biowaste during composting in a monitored compost bin. *J. Appl. Microbiol.* **94**, 127–137 (2003).
- Zureik, M. et al. Sensitisation to airborne moulds and severity of asthma: cross sectional study from European Community respiratory health survey. *Br. Med. J.* **325**, 411–414 (2002).
- Nierman, W. C. et al. Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl Acad. Sci. USA* **101**, 14246–14251 (2004).

10. Zhou, S. *et al.* Whole-genome shotgun optical mapping of *Rhodobacter sphaeroides* strain 2.4.1 and its use for whole-genome shotgun sequence assembly. *Genome Res.* **13**, 2142–2151 (2003).
11. Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
12. Kodzius, R. *et al.* Rapid identification of allergen-encoding cDNA clones by phage display and high-density arrays. *Comb. Chem. High Throughput Screen.* **6**, 147–154 (2003).
13. Weichel, M. *et al.* Immunoglobulin E-binding and skin test reactivity to hydrophobin HCh-1 from *Cladosporium herbarum*, the first allergenic cell wall component of fungi. *Clin. Exp. Allergy* **33**, 72–77 (2003).
14. Bozza, S. *et al.* Vaccination of mice against invasive aspergillosis with recombinant *Aspergillus* proteins and CpG oligodeoxynucleotides as adjuvants. *Microbes Infect.* **4**, 1281–1290 (2002).
15. Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).
16. Gardiner, D. M., Jarvis, R. S. & Howlett, B. J. The ABC transporter gene in the sirodesmin biosynthetic gene cluster of *Leptosphaeria maculans* is not essential for sirodesmin production but facilitates self-protection. *Fungal Genet. Biol.* **42**, 257–263 (2005).
17. Bok, J. W. & Keller, N. P. LaeA, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryot. Cell* **3**, 527–535 (2004).
18. Galagan, J. E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* doi:10.1038/nature04341 (this issue).
19. Machida, M. *et al.* Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* doi:10.1038/nature04300 (this issue).
20. Mousavi, S. A. & Robson, G. D. Oxidative and amphotericin-mediated cell death in the opportunistic pathogen *Aspergillus fumigatus* is associated with an apoptotic-like phenotype. *Microbiol.* **150**, 1937–1945 (2004).
21. Thrane, C., Kaufmann, U., Stummann, B. M. & Olsson, S. Activation of caspase-like activity and poly (ADP-ribose) polymerase degradation during sporulation in *Aspergillus nidulans*. *Fungal Genet. Biol.* **41**, 361–368 (2004).
22. Fontaine, T. *et al.* Molecular organization of the alkali-insoluble fraction of *Aspergillus fumigatus* cell wall. *J. Biol. Chem.* **275**, 27594–27607 (2000).
23. Klis, F. M., Mol, P., Hellingwerf, K. & Brul, S. Dynamics of cell wall structure in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **26**, 239–256 (2002).
24. Zegers, I., Martins, J. C., Willem, R., Wyns, L. & Messens, J. Arsenate reductase from *S. aureus* plasmid pI258 is a phosphatase drafted for redox duty. *Nature Struct. Biol.* **8**, 843–847 (2001).
25. Hayakawa, T., Kobayashi, Y., Cui, X. & Hirano, S. A new metabolic pathway of arsenite: arsenic–glutathione complexes are substrates for human arsenic methyltransferase Cyt19. *Arch. Toxicol.* **79**, 183–191 (2005).
26. Wysocki, R., Bobrowicz, P. & Ulaszewski, S. The *Saccharomyces cerevisiae* ACR3 gene encodes a putative membrane protein involved in arsenite transport. *J. Biol. Chem.* **272**, 30061–30066 (1997).
27. Bentley, R. & Chasteen, T. G. Microbial methylation of metalloids: arsenic, antimony, and bismuth. *Microbiol. Mol. Biol. Rev.* **66**, 250–271 (2002).
28. Blyth, W., Grant, I. W., Blackadder, E. S. & Greenberg, M. Fungal antigens as a source of sensitization and respiratory disease in Scottish maltworkers. *Clin. Allergy* **7**, 549–562 (1977).
29. Pain, A. *et al.* Insight into the genome of *Aspergillus fumigatus*: analysis of a 922 kb region encompassing the nitrate assimilation gene cluster. *Fungal Genet. Biol.* **41**, 443–453 (2004).
30. Kim, H. *et al.* Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses. *Biotechniques* **33**, 924–930 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Initial work was funded by the Fungal Research Trust and Burroughs Wellcome Fund. Major funding came from the National Institute of Allergy and Infectious Diseases (NIAID), the Wellcome Trust and the Fondo de Investigaciones Sanitarias. Construction of the Af293 microarray was funded by NIAID. Additional BAC end sequencing was funded internally by the Institut Pasteur. We thank D. Dixon, C. Caulcott, V. McGovern, P. Goodwin and J.-L. Rodriguez-Tudela for their support and encouragement during this project. We also thank C. Staben of the University of Kentucky for intellectual assistance and script development.

Author Information The genome sequence has been submitted to GenBank under the accession numbers NC_007194–NC_007201. All microarray expression data are available through ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>) with accession numbers A-MEXP-205 (array design) and E-MEXP-332 and E-MEXP-333 (experimental data). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.N. (wnierman@tigr.org).

Genome sequencing and analysis of *Aspergillus oryzae*

Masayuki Machida¹, Kiyoshi Asai², Motoaki Sano¹, Toshihiro Tanaka³, Toshitaka Kumagai², Goro Tera^{2,20}, Ken-Ichi Kusumoto⁴, Toshihide Arima⁵, Osamu Akita⁵, Yutaka Kashiwagi⁴, Keietsu Abe⁶, Katsuya Gomi⁶, Hiroyuki Horiuchi⁷, Katsuhiko Kitamoto⁷, Tetsuo Kobayashi⁸, Michio Takeuchi⁹, David W. Denning¹⁰, James E. Galagan¹¹, William C. Nierman^{12,13}, Jiujiang Yu¹⁴, David B. Archer¹⁵, Joan W. Bennett¹⁶, Deepak Bhatnagar¹⁴, Thomas E. Cleveland¹⁴, Natalie D. Fedorova¹², Osamu Gotoh², Hiroshi Horikawa³, Akira Hosoyama³, Masayuki Ichinomiya⁷, Rie Igarashi³, Kazuhiro Iwashita⁵, Praveen Rao Juvvadi⁷, Masashi Kato⁸, Yumiko Kato³, Taishin Kin², Akira Kokubun³, Hiroshi Maeda⁶, Noriko Maeyama³, Jun-ichi Maruyama⁷, Hideki Nagasaki², Tasuku Nakajima⁶, Ken Oda⁵, Kinya Okada², Ian Paulsen¹², Kazutoshi Sakamoto⁵, Toshihiko Sawano³, Mikio Takahashi³, Kumiko Takase¹, Yasunobu Terabayashi¹, Jennifer R. Wortman¹², Osamu Yamada⁵, Youhei Yamagata⁶, Hideharu Anazawa¹⁷, Yoji Hata¹⁸, Yoshinao Koide¹⁹, Takashi Komori²⁰, Yasuji Koyama²¹, Toshitaka Minetoki²², Sivasundaram Suharnan²³, Akimitsu Tanaka²⁴, Katsumi Isono³, Satoru Kuhara²⁵, Naotake Ogasawara²⁶ & Hisashi Kikuchi³

The genome of *Aspergillus oryzae*, a fungus important for the production of traditional fermented foods and beverages in Japan, has been sequenced. The ability to secrete large amounts of proteins and the development of a transformation system¹ have facilitated the use of *A. oryzae* in modern biotechnology^{2–4}. Although both *A. oryzae* and *Aspergillus flavus* belong to the section *Flavi* of the subgenus *Circumdati* of *Aspergillus*, *A. oryzae*, unlike *A. flavus*, does not produce aflatoxin, and its long history of use in the food industry has proved its safety. Here we show that the 37-megabase (Mb) genome of *A. oryzae* contains 12,074 genes and is expanded by 7–9 Mb in comparison with the genomes of *Aspergillus nidulans*⁵ and *Aspergillus fumigatus*⁶. Comparison of the three aspergilli species revealed the presence of syntenic blocks and *A. oryzae*-specific blocks (lacking synteny with *A. nidulans* and *A. fumigatus*) in a mosaic manner throughout the genome of *A. oryzae*. The blocks of *A. oryzae*-specific sequence are enriched for genes involved in metabolism, particularly those for the synthesis of secondary metabolites. Specific expansion of genes for secretory hydrolytic enzymes, amino acid metabolism and amino acid/sugar uptake transporters supports the idea that *A. oryzae* is an ideal microorganism for fermentation.

Sequencing of the *A. oryzae* genome was accomplished using the whole-genome shotgun (WGS) approach. The 37-Mb genome was predicted to contain a total of 12,074 genes encoding proteins with a length greater than 100 amino acid residues (see Methods). The genome was confirmed to comprise eight chromosomes

(chromosomes 1–8 in decreasing size), the assignment of which is different from a previous report⁷ (Supplementary Table S1 and schematic drawing in Supplementary Fig. S1). Interestingly, the *A. oryzae* genome contained numerous stretches (1,750) of (A+T)-rich sequence (that is, >90% A+T composition in 50 nucleotides or longer), 6–9 times more than for *A. fumigatus* (197) and *A. nidulans* (308).

The *A. oryzae* genome is larger than those of *A. fumigatus* and *A. nidulans* by approximately 34% and 29%, respectively. Syntenic analysis of the three aspergilli revealed the presence of syntenic blocks and *A. oryzae*-specific blocks of sequence (lacking synteny with the two other aspergilli) in a mosaic manner throughout the *A. oryzae* genome (Fig. 1). Phylogenetic analysis of the three aspergilli using the whole-genome data showed that *A. nidulans* branched off earlier than *A. oryzae* and *A. fumigatus*⁵. Thus, the increase in genome size seems to be due to an *A. oryzae* lineage-specific acquisition of sequence, rather than loss of sequence in *A. nidulans* and *A. fumigatus*. If, on the other hand, *A. nidulans* and *A. fumigatus* are assumed to have lost 7–9 Mb of sequence after branching off from their *A. oryzae*-like ancestor, a greater proportion of syntenic blocks would be conserved between each of them and *A. oryzae* than between the two. However, we observed an almost equal proportion of syntenic blocks in the three species. This suggests that the genome size differences are largely due to sequence acquisition in *A. oryzae*. The expansion in genome size appears to be characteristic of the organisms closely related to *A. oryzae*, as the estimated genome size of

¹Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Higashi 1-1-1, Tsukuba, Ibaraki 305-8566, Japan.

²Computational Biology Research Center, AIST, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan. ³National Institute of Technology and Evaluation, Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan. ⁴National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan. ⁵National Research Institute of Brewing, 3-7-1 Kagamiyama, Higashihiroshima, Hiroshima 739-0046, Japan. ⁶Tohoku University, 1-1 Tsutsumidori-Amamiyamachi, Aoba-ku, Sendai 981-8555, Japan. ⁷The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. ⁸Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan. ⁹Tokyo University of Agriculture and Technology, Saiwai-cho 3-5-8, Fuchu, Tokyo 183-0054, Japan. ¹⁰The University of Manchester, Manchester M23 9PL, UK. ¹¹Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02142, USA. ¹²The Institute for Genomic Research, Rockville, Maryland 20850, USA. ¹³The George Washington University School of Medicine, Department of Biochemistry and Molecular Biology, 2300 Eye Street NW, Washington DC 20037, USA. ¹⁴USDA/ARS Southern Regional Research Center, 1100 Robert E. Lee Boulevard, New Orleans, Louisiana 70124, USA. ¹⁵School of Biology, University of Nottingham, Nottingham NG7 2RD, UK. ¹⁶Tulane University, New Orleans, Louisiana 70118, USA. ¹⁷Kyowa Hakko Kogyo Co. Ltd, 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-8185, Japan. ¹⁸Research Institute, Gekkeikan Sake Co. Ltd, 24 Shimotoba-koyanagi-cho, Fushimi-ku, Kyoto 612-8361, Japan. ¹⁹Amano Enzyme Inc., 4-179-35, Sue-cho, Kakamigahara, Gifu 509-0108, Japan. ²⁰INTEC Web and Genome Informatics Corporation, 1-3-3 Shinsuna, Koto-ku, Tokyo 136-8637, Japan. ²¹Kikkoman Corporation, 399 Noda, Noda, Chiba 278-0037, Japan. ²²Ozeki Co. Ltd., 4-9 Imadudezaike-cho, Nishinomiya, Hyogo 663-8227, Japan. ²³Axiohelix, 2-45, Aomi, Koto-ku, Tokyo 135-0064, Japan. ²⁴Higeta Shoyu, Co. Ltd., 2-8 Chuo-cho, Choshi, Chiba 288-8680, Japan. ²⁵Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan. ²⁶Nara Institute of Science and Technology, 8916-5, Takayama, Ikoma, Nara 630-0101, Japan.

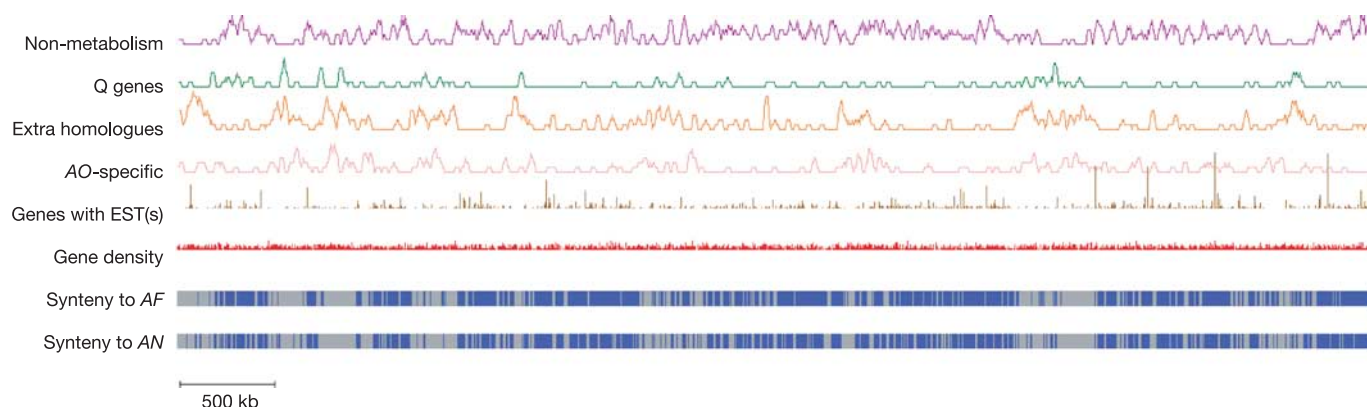


Figure 1 | Distribution and expression of the genes on chromosome 1. The blue bars at the bottom indicate the regions syntenic with *A. fumigatus* (AF) and *A. nidulans* (AN) genomes (see Methods). Non-metabolism, the genes relating to the COG categories other than metabolism; Q genes,

secondary metabolism genes; Extra homologues, extra *A. oryzae*-specific homologues; AO-specific, *A. oryzae*-specific genes; Genes with EST(s), genes that have one or more corresponding EST(s); Gene density, distribution of all the predicted genes. Synteny was analysed as described in the Methods.

its close relatives *A. flavus* (W. Nierman, personal communication) and *Aspergillus niger*⁸ is comparable to that of *A. oryzae*.

Using the cluster of orthologous group (COG)⁹ classification, most of the gene family expansion in the *A. oryzae* genome as compared to *A. fumigatus* was found to have occurred in those predicted to have roles in metabolism (C to Q), of which those for secondary metabolism (Q) are most significantly increased (Fig. 2; see also Supplementary Table S2). No significant differences were observed in the number of genes for any other COG category in comparison with *A. nidulans* and *A. fumigatus*, except for the genes involved in defence mechanisms (V) and extracellular structures (M).

These secondary metabolism genes are enriched in regions lacking synteny with either *A. fumigatus* or *A. nidulans* ($P = 9.8 \times 10^{-32}$, see Fig. 1 for chromosome 1 and Supplementary Fig. S2 for all eight chromosomes), and the genes having expressed sequence tags (ESTs) are considerably enriched in the syntenic regions ($P = 4.1 \times 10^{-134}$). Many more cytochrome P450 genes were observed in *A. oryzae* (149) compared with *A. nidulans* (102) and *A. fumigatus* (65) (Table 1). Of the polyketide synthase (PKS) genes, a specific expansion of WA-like PKS genes was observed (Supplementary Table S3). In addition, the *A. oryzae* genome contained a variety of homologues of trichothecene hydroxylases, isotrichodermin hydroxylases and tri-

chodiene oxygenases, as well as pisatin demethylases that are used by plant pathogenic fungi (for example, *Nectria haematococca*, *Fusarium* spp.) for detoxification of antimicrobial agents¹⁰. This is consistent with the close phylogenetic relationship of *A. oryzae* with the opportunistic plant pathogen *A. flavus*.

Although genes predicted to be involved in the aflatoxin synthetic pathway are present in *A. oryzae*, no ESTs of these genes were detected except for *aflJ* and *norA* (Akao, T. *et al.*, unpublished data), whereas ESTs for all 25 of the aflatoxin pathway genes were found in *A. flavus*¹¹. *A. oryzae* might have been selected as a non-toxigenic strain either during the long history of its industrial use or from the beginning.

In *A. oryzae*, all of the COG categories related to metabolism show an expansion of gene content (Fig. 2), the highest increase of which was observed for those involved in phenylalanine/tryptophan degradation (2 and 6 in Supplementary Fig. S3) and toluene/*m*-cresol/*p*-cymene degradation (9, 11 and 12 in Supplementary Fig. S3). This was based on the analysis using the *Saccharomyces cerevisiae* metabolic map as a reference. BAT1 and BAT2, which contribute to the metabolism of hydrophobic amino acids lysine and serine, are also over-represented (see Supplementary Fig. S4 for the entire metabolic pathways). There is also a significant expansion in the ATP-binding cassette (ABC), the amino acid-polyamine-organocation (APC) and the major facilitator

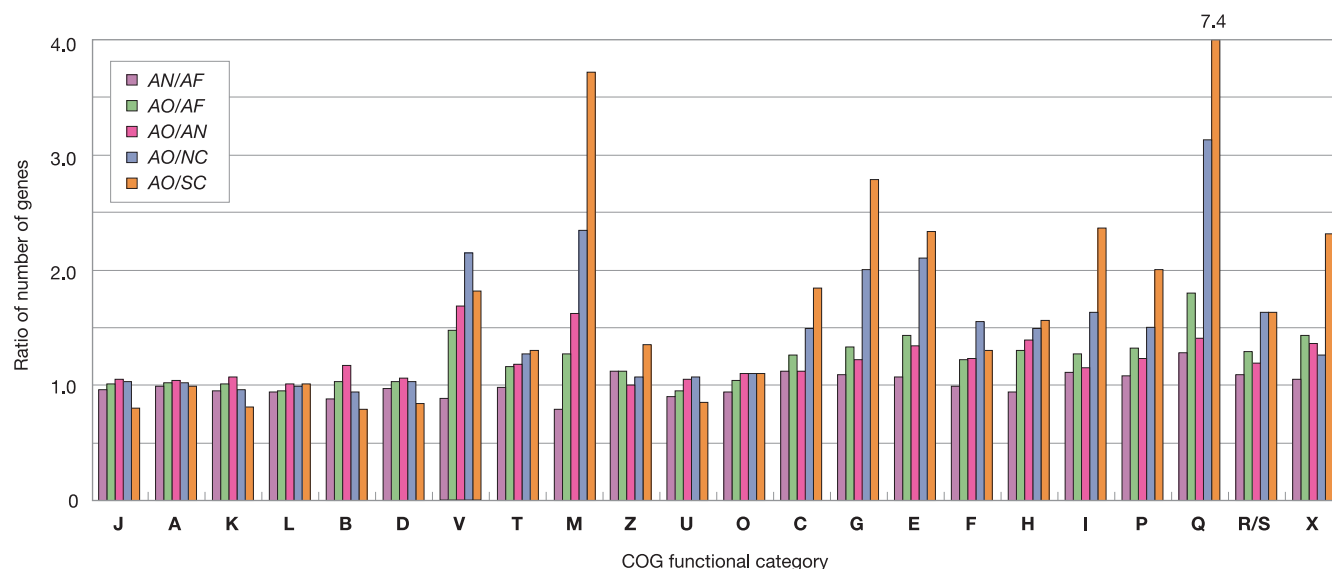


Figure 2 | Comparison of relative gene numbers for each COG. The ratios of the number of genes in *A. oryzae* against those in *A. fumigatus* (AO/AF), *A. nidulans* (AO/AN), *N. crassa* (AO/NC) and *S. cerevisiae* (AO/SC) for each COG category⁹ were calculated. The ratio of the number of genes in

A. nidulans against *A. fumigatus* (AN/AF) was also indicated. X indicates the genes without homology to any of the COG categories (see Methods). COGs with a gene number ≤ 5 for each species (Y, N and W) are not displayed to avoid misinterpretation derived from their possibly low reliability.

superfamily (MFS) transporter genes (Supplementary Table S4), which are concerned with multidrug resistance, transport of amino acids and transport of sugars, respectively.

Within the koji culture, *A. oryzae* grows on the surface of solid material such as steamed rice or ground soybean, where amino acids and sugars are deficient at the beginning. The need for *A. oryzae* to get access to external nitrogen sources effectively and to degrade proteins and starches seems consistent with the observed expansion of the metabolism and transporter-related gene families. Judging from the EST data, the genes for alcohol dehydrogenase, pyruvate decarboxylase and sugar transporters are typical examples of the *A. oryzae* genes that are transcribed most strongly (Akao, T. *et al.*, unpublished data). The strong expression of such genes might also have been enhanced through various adaptations¹² during the course of domestication.

Aspergilli possess more sensor histidine kinases (13–15) than *S. cerevisiae* (1) and *Schizosaccharomyces pombe* (3), whereas histidine-containing phosphotransfer factors and response regulators are found in similar numbers. *Aspergillus* histidine kinases are classified into nine families (HK1–9), of which the HK8 orthologue is absent in *Neurospora crassa* and the sequenced plant pathogens *Cochliobolus heterostrophus*, *Gibberella moniliformis*, *Fusarium graminearum* and *Magnaporthe grisea*. Whereas *A. fumigatus*, *A. nidulans*, *N. crassa* and the plant pathogens possess a single HK6 gene (*Nik-1* in *N. crassa*) that is essential for growth in high osmotic pressure, *A. oryzae* has two additional homologues. Continuous culturing under high osmolarity conditions (possibly through koji cultures) may have led to *A. oryzae* acquiring the additional *Nik-1* homologues. There are three MAPKKs and MAPKKs in the genomes of the three *Aspergillus* species and *N. crassa*. However, whereas *A. nidulans* and *A. fumigatus* possess four MAPKs and *A. oryzae* five, *N. crassa*, *F. graminearum* and *M. grisea* possess only three. Thus, *A. oryzae* may possess the most complex signal transduction cascade among the four filamentous fungi.

A. oryzae has the largest expansion of hydrolytic genes among the three aspergilli (Supplementary Table S5). The genomes of *A. oryzae*, *A. fumigatus* and *A. nidulans* contain 135, 99 and 90 secreted proteinase genes, respectively, which constitute roughly 1% of the total genes in each genome (Supplementary Table S6). All of the proteinase genes found in *A. fumigatus* and/or *A. nidulans* have orthologues in *A. oryzae* except for the one encoding aminopeptidase. On the other hand, several *A. oryzae* proteinase genes are missing in *A. fumigatus* and *A. nidulans*. Similarly, *A. oryzae* possesses more secretory proteinase genes that function in acidic pH, including aspartic proteinase, pepstatin-insensitive proteinase, serine type carboxypeptidase and aorsin (Supplementary Table S6). These increases may reflect *A. oryzae*'s adaptation to acidic pH during the course of its domestication.

The phylogenetic tree of secretory aspartic proteinases from the three aspergilli genomes (Fig. 3) shows six homologous clusters (yellow boxes) distributed on all chromosomes other than chromosome 7. Their features, including intron conservation, are similar to each other except for cluster 4, which shows the highest diversity. Each cluster contains four member genes (blue boxes), namely three orthologues from each *Aspergillus* species and an extra *A. oryzae*-specific homologue. All of the extra *A. oryzae* homologues are located in the *A. oryzae*-specific regions, whereas the orthologous clusters are located in the common regions, except for AO070319000053 of cluster 4. It is interesting to note that the clustering feature of the orthologues and extra homologues for aspartic proteinases is also conserved with the genes for carboxypeptidases (Supplementary Fig. S5a) and metalloproteinases. In contrast, the number of genes encoding intracellular enzymes (Supplementary Fig. S5b), including serine proteinases, is consistent in the three aspergilli. A similar expansion pattern was also observed for the genes for maltases (Supplementary Fig. S5c) and extracellular α -glucosidases. Besides the secretory hydrolases, some metabolic genes, including those in glucose fermentation and lysine biosynthesis, showed a similar gene expansion pattern (Supplementary Fig. S6).

It is well known that *A. oryzae* has three α -amylase genes (taka-amylase genes: *amyA*, *amyB* and *amyC*)¹³ that have almost identical nucleotide sequences with only one and two mismatches in the 5'-flanking and coding regions, respectively. The *amyA* gene has a transposon-like element at its 5'-flanking region, and the *amyB* and *amyC* genes have highly similar nucleotide sequences spanning approximately 5 kilobases (kb), including an incomplete transposon sequence at their 5'-flanking region. Phylogenetic analysis supports gene duplication to account for the expansion of the three α -amylase genes after *A. oryzae* branched off from the other two *Aspergillus* species (Supplementary Fig. S5d)—this is in clear contrast to the mode of gene expansion for the secretory proteinases mentioned above.

In contrast to the overall increase in the number of proteinases, *A. oryzae* has fewer glycosyl hydrolases with a cellulose-binding domain (five genes) or a starch-binding domain (*glaA*)¹⁴ to digest insoluble cellulose or raw and granular starch, respectively (Supplementary Table S5). Apparently, no additional enzymes for accessing carbohydrates are required during fermentation in contrast to those, including knottins, found in *A. fumigatus*, which seems appropriate for its ecological niche of rotting vegetable matter.

Protein folding in the endoplasmic reticulum is assisted by chaperones (for example, BiP, calnexin) and foldases (three protein-disulphide isomerase family proteins and a peptidyl-prolyl *cis-trans* isomerase). As in other fungi, however, there is no calreticulin homologue (Supplementary Table S7). Major secretory component genes, which alter the efficiency of protein secretion, were identified

Table 1 | Redundancy of the cytochrome P450 genes in aspergilli

Family*	Function	<i>A. oryzae</i>	<i>A. nidulans</i>	<i>A. fumigatus</i>
<i>cyp57</i>	Pisatin demethylases	12†	8†	3
<i>cyp58</i>	Trichodiene oxygenases	11†	9†	2
<i>cyp53</i>	Benzoate monooxygenases/hydroxylases	10†	5	6
<i>cyp64</i>	P450 oxidoreductases	10†	4	4
<i>cyp65</i>	Trichothecene hydroxylases	9†	3	1
<i>cyp52</i>	Alkane hydroxylases	8	6	8
<i>cyp65</i>	Isotrichodermin hydroxylases	5†	3†	1
<i>cyp505</i>	Fatty acid hydroxylases	4†	2	2
<i>cyp51</i>	Sterol demethylases	3	2	2
<i>cyp509</i>	Fum15 homologues	3	2	2
<i>cyp61</i>	Sterol desaturases	2†	1	1
<i>cyp55</i>	P450 nitric oxide reductase	1	–	–
<i>cyp58</i> , <i>cyp59</i> , <i>cyp60</i> , <i>cyp62</i> , <i>cyp68</i> , <i>cyp53</i> , <i>cyp503</i> , <i>cyp512</i>	P450 monooxygenases	22†	20†	7
Unknown cytochrome P450s		49	37	25
Total cytochrome P450s		149†	102	65

*Classification of the genes was performed using the P450 Blast server (<http://drnelson.utmem.edu/CytochromeP450.html>) according to the P450 nomenclature conventions.

†The number of genes is $\geq$ twofold of the minimum number among the three aspergilli.

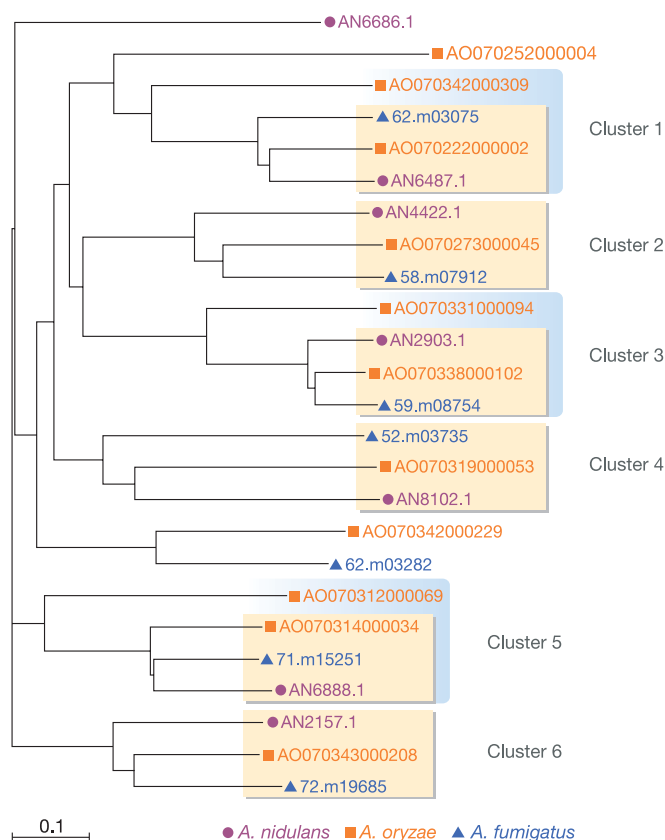


Figure 3 | Phylogenetic analysis of aspartic proteinases. The phylogenetic relationship of aspartic proteinase homologues from the three *Aspergillus* was analysed by the ClustalX³⁰ program, successive unweighted pair-group method using arithmetic averages (UPGMA), and drawn by TreeView (Roderic, D. M., <http://taxonomy.zoology.gla.ac.uk/rod/rod.html>). Orange, blue and purple characters designate the *A. oryzae*, *A. fumigatus* and *A. nidulans* genes, respectively. Orthologous clusters among the three *Aspergillus* and the clusters with an extra *A. oryzae* homologue are indicated by yellow and blue boxes, respectively.

in all three *Aspergillus* with an exception of the *A. fumigatus* SSS1 homologue (Supplementary Table S8).

In comparison to the common regions, the *A. oryzae*-specific regions contained 1.7 times lower density of genes homologous to those in eukaryotes other than *A. fumigatus* and *A. nidulans*. In a search for bacterial homologues, we found two genes (AO070319000101 and AO070319000102) in an *A. oryzae*-specific region with highest sequence similarity to those of *Agrobacterium tumefaciens* (AGR_L_1864 (biotin carboxylase) and AGR_L_1866, hypothetical protein genes with *E*-values of 0.0 and 1×10^{-119} , respectively). Because the two genes are adjacently located in both *A. oryzae* and *A. tumefaciens* (Supplementary Fig. S7a), and the two *A. oryzae* genes reside in a 'bacterial cluster' (Supplementary Fig. S7b), they are suggested to have been laterally transferred.

The expansion of *A. oryzae*-specific homologues might be the result of genome-wide duplication, as observed in yeast. The speciation of *Aspergillus* was estimated to have taken place approximately 20 million years ago¹⁵ and was later than the whole-genome duplication event in yeast, which was estimated to have taken place 150 million years ago¹⁶. We were unable to observe any extended stretch of region within the *A. oryzae* genome that showed a certain degree of similarity to another stretch of region despite the fact that we observed synteny among the three *Aspergillus* (Fig. 1) and that segmentally duplicated stretches were detected by the same method within the *S. cerevisiae* genome. Thus, the increase in the genome size of *A. oryzae* relative to *A. fumigatus* and *A. nidulans* does not appear to be due to chromosomal duplication. The large segmental

duplication, if any, must have taken place much earlier than the separation of the three *Aspergillus*, and the similarity between the duplicated regions might have been completely lost by extensive sequence alterations and rearrangements. However, if the three *Aspergillus* had a common ancestor possessing the expanded gene families found in *A. oryzae*, both *A. nidulans* and *A. fumigatus* must independently have lost approximately 3,000 genes in common with the putative common ancestor.

The mosaic structure of the genome, considered to be evidence for horizontal gene transfer¹⁷, was found by synteny analysis of the *A. oryzae* genome and was further characterized by the localization of the EST expression (see above) of non-metabolic genes ($P = 1.78 \times 10^{-95}$ and 1.32×10^{-51} for information and storage (J to B) and cellular function and signalling (D to O), respectively) and the genes of high codon adaptation index (top 5% genes, $P = 9.8 \times 10^{-28}$). The phylogenetic distance between the genes in the orthologous cluster and the *A. oryzae*-specific ones was similar to that between the genes of *Aspergillus* and the other genera belonging to Sordariomycetes. The statistical analysis by ref. 18 of some *A. oryzae*-specific homologues of aspartic proteinase, carboxypeptidase, maltase, pyruvate decarboxylase and lysine-ketoglutarate reductase/saccharopine dehydrogenase showed *P*-values of between 0.000 and 0.004. The results indicated phylogenetic inconsistency of these genes. These results, together with the above discussion, imply that the *A. oryzae*-specific genes have been transferred by a similar mechanism observed for an asexual pathogenic fungus¹⁹, in which chromosomes are transferred between genetically isolated clonal lines. It has been reported that yeast chromosomes are rearranged frequently under starved culture conditions and that (A+T)-rich sequences or transfer RNA often mediate such rearrangements²⁰. Our EST analysis shows that the expression profile in solid-state cultivation is similar to that observed when a carbon source is omitted²¹. These reports suggest that the acquired foreign DNA has been rearranged in a short period of time by large-scale solid-state cultivation since *A. oryzae* was domesticated from an ancestor of *A. flavus*²².

It is tempting to speculate that the gene expansion of *A. oryzae* is explained by horizontal gene transfer; however, at this moment we cannot exclude the possibility of massive gene loss in the two other *Aspergillus* species. Future comparative analyses with more closely related species would provide more insight into the scenario of the genome evolution of *A. oryzae*, including that which occurred during the centuries of domestic cultivation.

METHODS

Strain and DNA preparation. *Aspergillus oryzae* RIB40 (National Research Institute of Brewing Stock Culture and ATCC-42149) was used as the DNA donor. Genomic DNA preparation and removal of mitochondrial DNA was performed as described by refs 23 and 24, respectively.

Genome sequencing. The genome sequencing of *A. oryzae* was accomplished using the WGS approach by accumulating raw sequence reads of approximately $\times 9$ depth of coverage. Contigs generated were mapped by Southern hybridization onto chromosomes separated by PFGE. Linkage between contigs was analysed by fingerprinting and PCR methods. Sequence assembly was validated with high-density end sequences of bacterial artificial chromosome (BAC) and cosmid clones and by Optical Mapping (OpGen). See Supplementary Information for details.

Gene prediction and annotation. Genes were predicted in the *A. oryzae* genome based on the homologies to known genes in the public database, ESTs of *A. oryzae* and *A. flavus*, and the statistical features of the genes by applying a combination of gene-finding software. Transfer RNAs were identified using tRNAscan-SE²⁵. Repeated sequences were detected using RepeatMasker (Smit, A. F. A. and Green, P., <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). The homologues of the proteins of *Aspergillus*, *N. crassa*, *M. grisea*, *Gibberella zeae*, *Penicillium* and *Paecilomyces* are searched for by running BlastX with a threshold value of $E \leq 1 \times 10^{-10}$. The resultant candidates of homologues are evaluated by ALN²⁶, which predicts the precise gene structures by aligning the Blast hits and the protein sequences. ALN takes into account frameshift errors, coding potentials and signals for translational initiation, termination and splicing. Of

the 6,586 genes thus predicted by ALN, 489 highly reliable genes were adopted into a learning set for GeneDecoder²⁷ and GlimmerM²⁸ software that work based on the statistic features of genes. GeneDecoder also integrates the information for splice sites provided by the ESTs, which are aligned with the genome sequence by SIM4 (ref. 29). Fivefold cross-validation of the gene finders trained by the above data set showed sensitivity/specificity for the exon prediction of 0.74/0.53 and 0.66/0.59 for GeneDecoder and GlimmerM, respectively, and those for coding sequences of 0.93/0.90 and 0.92/0.98. Genes partially supported by ESTs were predicted by GeneDecoder and those without any support by the known genes or ESTs were predicted by GlimmerM. The numbers of genes predicted by ALN, GlimmerM and GeneDecoder were 5,367, 6,983 and 1,713, respectively. All of the predicted protein-coding genes were annotated by searching against the COG database⁹ using BlastP, followed by manual corrections.

Syntenic analysis. Orthologues between *A. oryzae* and either *A. nidulans* or *A. fumigatus* were identified using the best bi-directional hit method (BlastP with a bit score greater than 200). In addition, putative homologous regions between the species were identified by TblastX with a bit score greater than 100. Orthologues and homologous regions between the contigs of two species were aligned to make a contiguous block, until no orthologues or homologous regions were found within the range of 10 kb. A region of conserved synteny was defined as the longest contiguous block that contained at least one orthologue and one additional orthologue or homologous region.

COG analysis. The number of genes for each COG category was analysed by a BlastP search using the amino acid sequences in the COG set⁹ with the bit score of ≥ 60 .

Gene localization. Distribution of all predicted genes and the genes with ESTs that were obtained from mycelia grown in either liquid-rich medium, liquid-starved medium or solid-state cultivation (Akao, T. *et al.*, unpublished data) were analysed by counting the corresponding genes in a 5-kb window. Distributions of non-metabolic genes, secondary metabolism genes, extra *A. oryzae*-specific homologues that have homology (bit score ≥ 100) to orthologues identified by best bi-directional match between *A. oryzae* and either *A. fumigatus* or *A. nidulans*, as well as *A. oryzae*-specific genes without homology to either *A. fumigatus* or *A. nidulans* genes (bit score < 100) were analysed in the same way by applying a window size of 15 kb.

Statistical analyses. Localization of the secondary metabolism genes at the *A. oryzae*-specific regions was evaluated by the one-tailed *P*-value based on the binomial distribution with the sample size of 413. Localization of the genes with EST expression, non-metabolic genes and the top 5% of genes with a high CAI value at the syntenic regions was evaluated in the same way with sample sizes of 33,77, 1,839 and 703, respectively. The analyses were performed when *A. oryzae*-specific regions were detected by comparing the *A. oryzae* and *A. fumigatus* genomes. The phylogenetic inconsistency was statistically analysed by the method described in ref. 18 using data sets consisting of the genes of the three aspergilli and three species belonging to Sordariomycetes or Eurotiomycetes other than *Aspergillus*. The reference and test data sets included the *A. oryzae* gene in the orthologous cluster and the extra *A. oryzae*-specific homologue, respectively.

Received 18 May; accepted 6 October 2005.

- Gomi, K., Iimura, Y. & Hara, S. Integrative transformation of *Aspergillus oryzae* with a plasmid containing the *Aspergillus nidulans* *argB* gene. *Agric. Biol. Chem.* **51**, 2549–2555 (1987).
- Christensen, T. *et al.* High level expression of recombinant genes in *Aspergillus oryzae*. *Bio/Technology* **6**, 1419–1422 (1988).
- Ward, P. P. *et al.* Production of biologically active recombinant human lactoferrin in *Aspergillus oryzae*. *Bio/Technology* **10**, 784–789 (1992).
- Tsuchiya, K. *et al.* High level expression of the synthetic human lysozyme gene in *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* **38**, 109–114 (1992).
- Galagan, J. E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* doi:10.1038/nature04341 (this issue).
- Nierman, W. *et al.* Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* doi:10.1038/nature04332 (this issue).
- Kitamoto, K., Kimura, K., Gomi, K. & Kumagai, C. Electrophoretic karyotype and gene assignment to chromosomes of *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.* **58**, 1467–1470 (1994).
- Archer, D. B. & Dyer, P. S. From genomics to post-genomics in *Aspergillus*. *Curr. Opin. Microbiol.* **7**, 499–504 (2004).
- Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).
- van den Brink, H. M., van Gorcom, R. F., van den Hondel, C. A. & Punt, P. J. Cytochrome P450 enzyme systems in fungi. *Fungal Genet. Biol.* **23**, 1–17 (1998).
- Yu, J., Whitelaw, C. A., Nierman, W. C., Bhatnagar, D. & Cleveland, T. E. *Aspergillus flavus* expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops. *FEMS Microbiol. Lett.* **237**, 333–340 (2004).
- Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl Acad. Sci. USA* **96**, 9721–9726 (1999).
- Sakaguchi, K., Takagi, M., Horiuchi, H. & Gomi, K. in *Applied Molecular Genetics in Filamentous Fungi* (eds Kinghorn, J. R. & Turner, G.) 54–99 (Blackie Academic & Professional, Glasgow, 1992).
- Hata, Y. *et al.* Nucleotide sequence and expression of the glucoamylase-encoding gene (*glaA*) from *Aspergillus oryzae*. *Gene* **108**, 145–150 (1991).
- Hori, H. & Osawa, S. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Mol. Biol. Evol.* **4**, 445–472 (1987).
- Langkjaer, R. B., Clifton, P. F., Johnston, M. & Piskur, J. Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. *Nature* **421**, 848–852 (2003).
- Prade, R. A., Griffith, J., Kochut, K., Arnold, J. & Timberlake, W. E. *In vitro* reconstruction of the *Aspergillus* (= *Emericella*) *nidulans* genome. *Proc. Natl Acad. Sci. USA* **94**, 14564–14569 (1997).
- Lawrence, J. G. & Hartl, D. L. Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. *Genetics* **131**, 753–760 (1992).
- Masel, A. M., He, C., Poplawski, A. M., Irwin, J. A. G. & Manners, J. M. Molecular evidence for chromosome transfer between biotypes of *Colletotrichum gloeosporioides*. *Mol. Plant–Microbe Interact.* **9**, 339–348 (1996).
- Dunham, M. J. *et al.* Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **99**, 16144–16149 (2002).
- Maeda, H. *et al.* Transcriptional analysis of genes for energy catabolism and hydrolytic enzymes in the filamentous fungus *Aspergillus oryzae* using cDNA microarrays and expressed sequence tags. *Appl. Microbiol. Biotechnol.* **65**, 74–83 (2004).
- Geiser, D. M., Pitt, J. I. & Taylor, J. W. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proc. Natl Acad. Sci. USA* **95**, 388–393 (1998).
- Iimura, Y., Gomi, K., Uzu, H. & Hara, S. Transformation of *Aspergillus oryzae* through plasmid-mediated complementation of the methionine-auxotrophic mutation. *Agric. Biol. Chem.* **51**, 323–328 (1987).
- Watson, J. & Thompson, W. F. Purification and restriction endonuclease analysis of plant nuclear DNA. *Methods Enzymol.* **118**, 57–75 (1986).
- Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
- Gotoh, O. Homology-based gene structure prediction: simplified matching algorithm using a translated codon (tron) and improved accuracy by allowing for long gaps. *Bioinformatics* **16**, 190–202 (2000).
- Asai, K., Itou, K., Ueno, Y. & Yada, T. Recognition of human genes by stochastic parsing. *Pac. Symp. Biocomput.* **3**, 228–239 (1998).
- Majoros, W. H., Pertea, M., Antonescu, C. & Salzberg, S. L. GlimmerM, Economy and Unveil: three *ab initio* eukaryotic gene finders. *Nucleic Acids Res.* **31**, 3601–3604 (2003).
- Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M. & Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**, 967–974 (1998).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to N. Hall and H. Hagiwara for discussions and critical reading of the manuscript. We thank M. Tadenuma and T. Ishikawa of the Brewing Society of Japan for the office work necessary for the collaborative research work of companies, national institutes and universities.

Author Contributions M.S., T.T., K. Kusumoto, T.A., Y. Kashiwagi, H. Horikawa, A.H., R.I., Y. Kato, A.K., N.M., T.S., K.T., S.S., K. Isono, S.K., N.O., H.K. and M.M. sequenced the *A. oryzae* genome; the genes were computationally predicted and annotated from the *A. oryzae* genome by K.A., T. Kumagai, G.T., J.Y., D.B., T.E.C., O.G., T. Kin, H.N. and T. Komori; M.S., K. Kusumoto, T.A., O.A., Y. Kashiwagi, K.A., K.G., H. Horiuchi, K. Kitamoto, T. Kobayashi, M. Takeuchi, D.W.D., D.A., J.W.B., M.I., K. Iwashita, P.R.J., M.K., H.M., J.M., T.N., K. Oda, I.P., K.S., Y.T., O.Y., Y.Y., H.A., Y.H., Y. Koide, Y. Koyama, T.M., A.T. and M.M. contributed correction and hand annotation of the predicted genes; K.A., T. Kumagai, G.T., D.W.D., J.E.G., W.C.N., N.D.F., T. Kin, H.N., Y.T., J.W., T. Komori and M.M. analysed gene localization and development of the *A. oryzae* genome.

Author Information The genome sequence has been submitted to DDBJ under the accession numbers AP007150–AP007177. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (m.machida@aist.go.jp).

LETTERS

NMDA receptors are expressed in oligodendrocytes and activated in ischaemia

Ragnhildur Káradóttir¹, Pauline Cavelier¹, Linda H. Bergersen² & David Attwell¹

Glutamate-mediated damage to oligodendrocytes contributes to mental or physical impairment in periventricular leukomalacia (pre- or perinatal white matter injury leading to cerebral palsy), spinal cord injury, multiple sclerosis and stroke^{1–4}. Unlike neurons⁵, white matter oligodendrocytes reportedly lack NMDA (*N*-methyl-*D*-aspartate) receptors^{6,7}. It is believed that glutamate damages oligodendrocytes, especially their precursor cells, by acting on calcium-permeable AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)/kainate receptors alone^{1–4} or by reversing cystine–glutamate exchange and depriving cells of antioxidant protection⁸. Here we show that precursor, immature and mature oligodendrocytes in the white matter of the cerebellum and corpus callosum exhibit NMDA-evoked currents, mediated by receptors that are blocked only weakly by Mg^{2+} and that may contain NR1, NR2C and NR3 NMDA receptor subunits. NMDA receptors are present in the myelinating processes of oligodendrocytes, where the small intracellular space could lead to a large rise in intracellular ion concentration in response to NMDA receptor activation. Simulating ischaemia led to development of an inward current in oligodendrocytes, which was partly mediated by NMDA receptors. These results point to NMDA receptors of unusual subunit composition as a potential therapeutic target for preventing white matter damage in a variety of diseases.

Blocking AMPA/kainate receptors attenuates white matter injury in animal models of hypoxia/ischaemia^{9,10}, spinal cord injury^{11,12} and multiple sclerosis^{13,14}. However, NMDA receptors have been found in some cultured oligodendrocyte precursors¹⁵ and in spinal grey (but not white) matter oligodendrocytes¹⁶, and NMDA receptor blockers have been shown to slow the loss of action potentials in white matter¹⁰ and reduce damage to white matter in ischaemia¹⁷ and in a model of multiple sclerosis¹⁸. We have therefore re-examined the involvement of NMDA receptors in the physiology and pathology of oligodendrocytes.

Whole-cell clamped precursor, immature and mature oligodendrocytes in white matter were distinguished by their morphology and antibody labelling^{7,19} (Fig. 1a–c). Precursor cells (Fig. 1a) had short processes not aligned with nearby axons, were labelled by antibodies against NG2 proteoglycan (see Methods for details regarding specificity), and often showed spontaneous synaptic currents (similar to grey matter oligodendrocyte precursors²⁰) that were blocked by tetrodotoxin (TTX; Fig. 1d). Immature cells (Fig. 1b) had some processes aligned with adjacent axons, and were labelled by an antibody against O4 lipid sulphatide. Mature cells (Fig. 1c) had most of their processes aligned with axons, and were labelled by an antibody against myelin basic protein (MBP). Oligodendrocyte membrane resistance decreased with maturity (Fig. 1e). The resting potential was approximately -60 mV (-57.5 ± 2.9 mV in 37 mature cells; without shunting by the electrode seal it would be ~ 4 mV more negative).

Precursor, immature and mature cerebellar white matter oligodendrocytes responded to glutamate ($100 \mu\text{M}$) with an inward current at -63 mV, which was unaffected by application of $1 \mu\text{M}$ TTX ($P = 0.4$; Fig. 1f–h and Supplementary Fig. 1), showing that axonal action potentials generated by glutamate depolarizing neurons did not contribute to the current (for example, by releasing K^+). The current was potentiated when the local glutamate concentration was raised by blocking glutamate transporters with TBOA (threo- β -benzyloxyaspartate, $P = 0.016$; Fig. 1g and Supplementary Fig. 1). The NMDA receptor antagonists *D*-AP5 (*D*(-)-2-amino-5-phosphonovaleic acid) and MK801, and the AMPA receptor antagonist NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[*f*]quinoxaline-7-sulphonamide) all reduced the glutamate-evoked current (Fig. 1h, i), suggesting the presence of both NMDA and AMPA/kainate receptors.

NMDA ($60 \mu\text{M}$) evoked an inward current in corpus callosum (Fig. 2a–c) and cerebellum (Fig. 2d, e) oligodendrocytes that was comparable in size (at -63 mV in 0 mM Mg^{2+}) to the current produced by AMPA ($20 \mu\text{M}$) or kainate ($30 \mu\text{M}$), and was larger than that produced by the metabotropic glutamate receptor agonist (1*S*,3*R*)-ACPD ((1*S*,3*R*)-1-aminocyclopentane-1,3-dicarboxylic acid; $100 \mu\text{M}$). In cerebellar oligodendrocytes, the NMDA-evoked current was largest in mature cells (Fig. 2f, $P = 0.042$ compared with precursors) although the increased membrane area of mature cells may imply a lower current density.

The NMDA-evoked current ran down slightly with time (Fig. 2g), so current measurements in the presence of drug treatments were quantified relative to the average of control responses before and after drug application. The NMDA-evoked current was blocked by *D*-AP5 (Fig. 2h, j) but unaffected by TTX, NBQX or strychnine + bicuculline (Fig. 2i, j), and was not significantly affected ($P = 0.072$) when glycine ($100 \mu\text{M}$, with $5 \mu\text{M}$ strychnine) was present, implying that the glycine-binding sites on these NMDA receptors are well-activated by endogenous glycine or *D*-serine. Changing from Mg^{2+} -free superfusion solution to solution containing 2 mM Mg^{2+} decreased the NMDA-evoked current at -63 mV three- to fivefold, independent of developmental age (Fig. 2k, l). This decrease is much less than is found for most neurons²¹ or for most cloned NMDA receptors²², which show a 60-fold reduction for receptors comprising NR1 and NR2A or NR2B subunits, and a 20-fold reduction for receptors comprising NR1 and NR2C or NR2D subunits, but is comparable to that seen for cloned receptors comprising NR1, NR2A and NR3A subunits²³.

Ifenprodil ($10 \mu\text{M}$), which blocks NR2B-containing NMDA receptors, had no effect on the NMDA-evoked current (Fig. 2j). The current was also unaffected by pregnenolone sulphate ($100 \mu\text{M}$, which potentiates NR1/NR2A and NR1/NR2B receptors by 60–82% but inhibits NR1/NR2C and NR1/NR2D receptors by 32%), *D*-cycloserine (1 mM, with no added glycine, which potentiates NR1/NR2C

¹Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK. ²Department of Anatomy and Centre for Molecular Biology and Neuroscience, University of Oslo, PO Box 1105, Blindern, N-0317 Oslo, Norway.

receptors but inhibits NR1/NR2A and NR1/NR2B receptors) or D-serine (100 μ M, with no glycine, which potentiates receptors lacking NR3 subunits but inhibits NR3-containing receptors) (Fig. 2j). These results suggest either (1) the presence of a subunit (perhaps NR3) that suppresses modulatory actions on NR2 subunits, or (2) the presence of a combination of subunits that these agents modulate in opposite directions (for example, NR1/NR2A/NR2C or NR1/NR2/NR3), or (3) a mixture of receptors with different subunit combinations. Although the NMDA receptor subunit composition currently remains uncertain, it allows these receptors to generate a significant current even at the resting potential in a physiological extracellular Mg^{2+} concentration.

In precursor oligodendrocytes, the NMDA-evoked current reversed around 0 mV (Fig. 2m), and in 2 mM Mg^{2+} the current–voltage (I – V) relationship showed a region of negative slope that was similar to, but less marked than, that produced by Mg^{2+} -block of neuronal NMDA receptor channels. In mature oligodendrocytes, the I – V relationship often failed to reverse at positive potentials, which might reflect the NMDA receptors being electrotonically distant from the soma in the cell processes (see below), or NMDA receptor activation leading to Na^+ entry and block of a K^+ current in the cell²⁴.

Antibodies directed against NMDA receptor subunits labelled the myelinating processes, and some cell bodies, of oligodendrocytes in the cerebellar white matter. NR2C labelling (Fig. 3a, b) was abolished by omitting the primary antibody, or by preabsorption with the

peptide to which it was raised (Supplementary Fig. 2). Labelling was also seen for NR1 subunits, which colocalized with MBP in mature cells (data not shown), and for NR3 subunits (Fig. 3c, d and Supplementary Fig. 2). Weaker labelling was seen for NR2A, NR2B and NR2D, and was abolished by omitting the primary antibody (NR2A, NR2B and NR2D) or by peptide absorption (NR2D). Double-labelling showed colocalization of NR3 and NR2C subunits, and of NR1 and NR2C subunits in oligodendrocyte processes (Supplementary Figs 3, 4). These data, together with the lack of ifenprodil-induced block and the weak Mg^{2+} -block of the NMDA-evoked current, suggest that oligodendrocyte NMDA receptors contain at least NR1, NR2C and NR3 subunits.

Post-embedding electron microscopic immunocytochemistry showed that NR1 subunits were present in the myelinating processes of adult cerebellar oligodendrocytes (Fig. 3e, f and Supplementary Fig. 5), in the outer- and innermost membranes and also within the myelin (perhaps remaining from earlier in development). By quantifying the immunogold particles in the myelin and in the postsynaptic densities of mossy fibre–granule cell synapses (Supplementary Fig. 5) and parallel fibre–Purkinje cell synapses, we found that the density of NMDA receptors throughout the myelin is as high as at the mossy fibre–granule cell synapse (Fig. 3g). The density of particles in the outer membrane of the myelin (where the receptors may sense glutamate released from surrounding cells; 36.4 ± 8.3 particles per μm^2) tended to be larger than that in the innermost membrane (where the receptors may sense glutamate released from the axon;

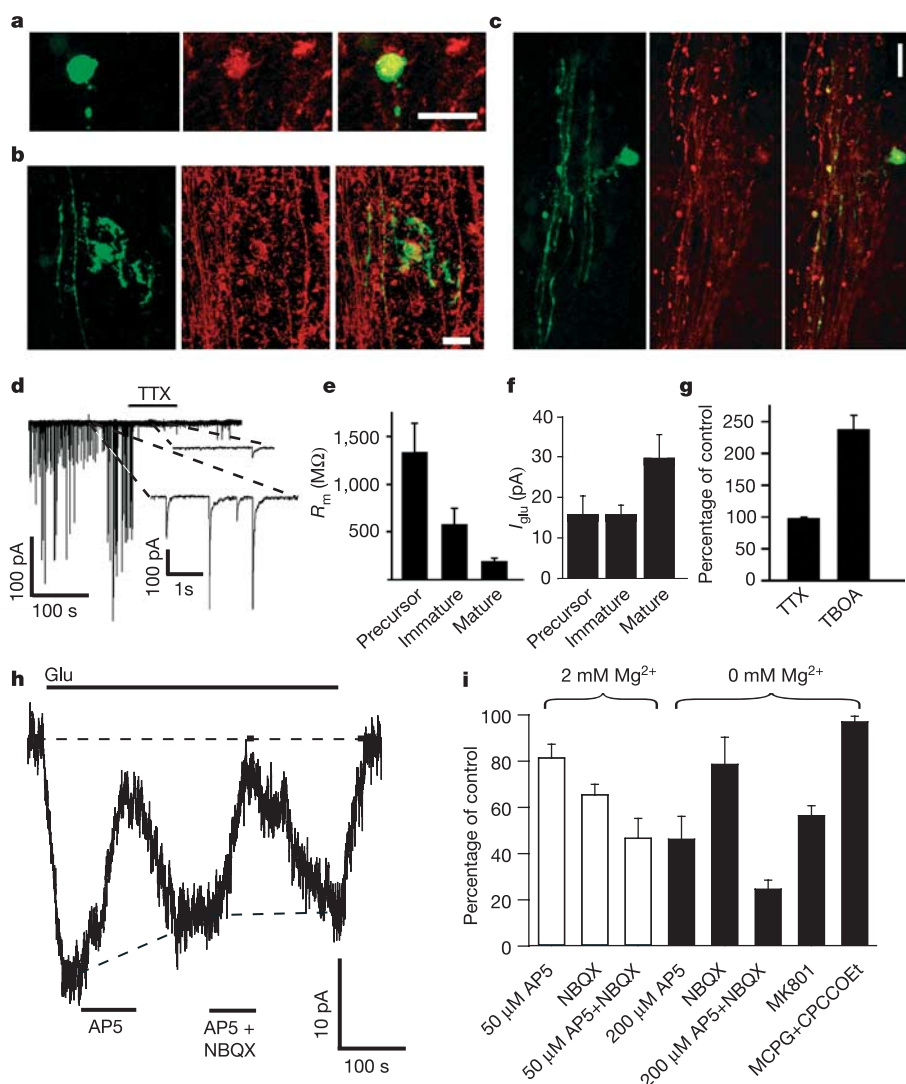


Figure 1 | Glutamate-evoked current in oligodendrocytes. **a–c**, Lucifer staining (green) and labelling with antibodies (red) against NG2 in precursor oligodendrocytes (**a**), O4 in immature oligodendrocytes (**b**) and MBP in mature oligodendrocytes (**c**). Colocalization shown as yellow in overlay. Scale bar, 20 μ m. **d**, TTX (1 μ M) blocks synaptic currents in a precursor oligodendrocyte. **e**, Membrane resistance ($\pm$ s.e.m.); $P < 0.001$, comparing precursors with immature or mature oligodendrocytes. **f**, Current evoked by 100 μ M glutamate in the presence of 2 mM Mg^{2+} ($P = 0.11$ and $P = 0.14$ comparing 33 mature oligodendrocytes with 22 immature or 22 precursor cells, respectively). **g**, Effect of TTX ($n = 5$ cells) and TBOA (200 μ M; $n = 4$) on glutamate-evoked current. **h**, Effect of 200 μ M D-AP5 (which blocks NMDA receptor responses by $\sim 78\%$, see Supplementary Material) and 25 μ M NBQX (which blocks AMPA responses by $>99\%$) on glutamate-evoked current in Mg^{2+} -free solution. **i**, Current remaining as a percentage of control in 2 mM Mg^{2+} solution ($n = 7$ cells) containing 50 μ M D-AP5 ($P = 0.02$, blocks NMDA receptors by $\sim 39\%$), 25 μ M NBQX ($P = 0.0003$) or AP5 + NBQX ($P = 0.0008$), and in Mg^{2+} -free solution ($n = 4–6$) containing 200 μ M AP5 ($P = 0.01$), 25 μ M NBQX ($P = 0.16$), AP5 + NBQX ($P = 0.0003$), 10 μ M MK801 ($P = 0.002$) or the metabotropic glutamate receptor blockers MCPG (1 mM) plus CPCCOEt (200 μ M; $P = 0.25$). **d–g**, In the presence of 2 mM Mg^{2+} ; **d–i**, -63 mV, 24° C. Data are presented as mean $\pm$ s.e.m.

17.7 ± 4.3 particles per μm^2 , $P = 0.064$), but both were similar to the average density measured over all of the myelin (33.6 ± 4.5 particles per μm^2).

To examine the role of oligodendrocyte NMDA receptors in pathological situations, we simulated²⁵ the energy deprivation that contributes to periventricular leukomalacia or that occurs after

ischaemia in spinal cord injury. A solution simulating ischaemia evoked a slowly developing inward current in precursor and mature oligodendrocytes (Fig. 4, peak inward current at -63 mV after ~ 7 min was 307 ± 83 pA and 280 ± 48 pA in 7 precursor and 8 mature cells, respectively, in 2 mM Mg^{2+}). This current was generated partly by an increase in extracellular glutamate concentration, as it was reduced

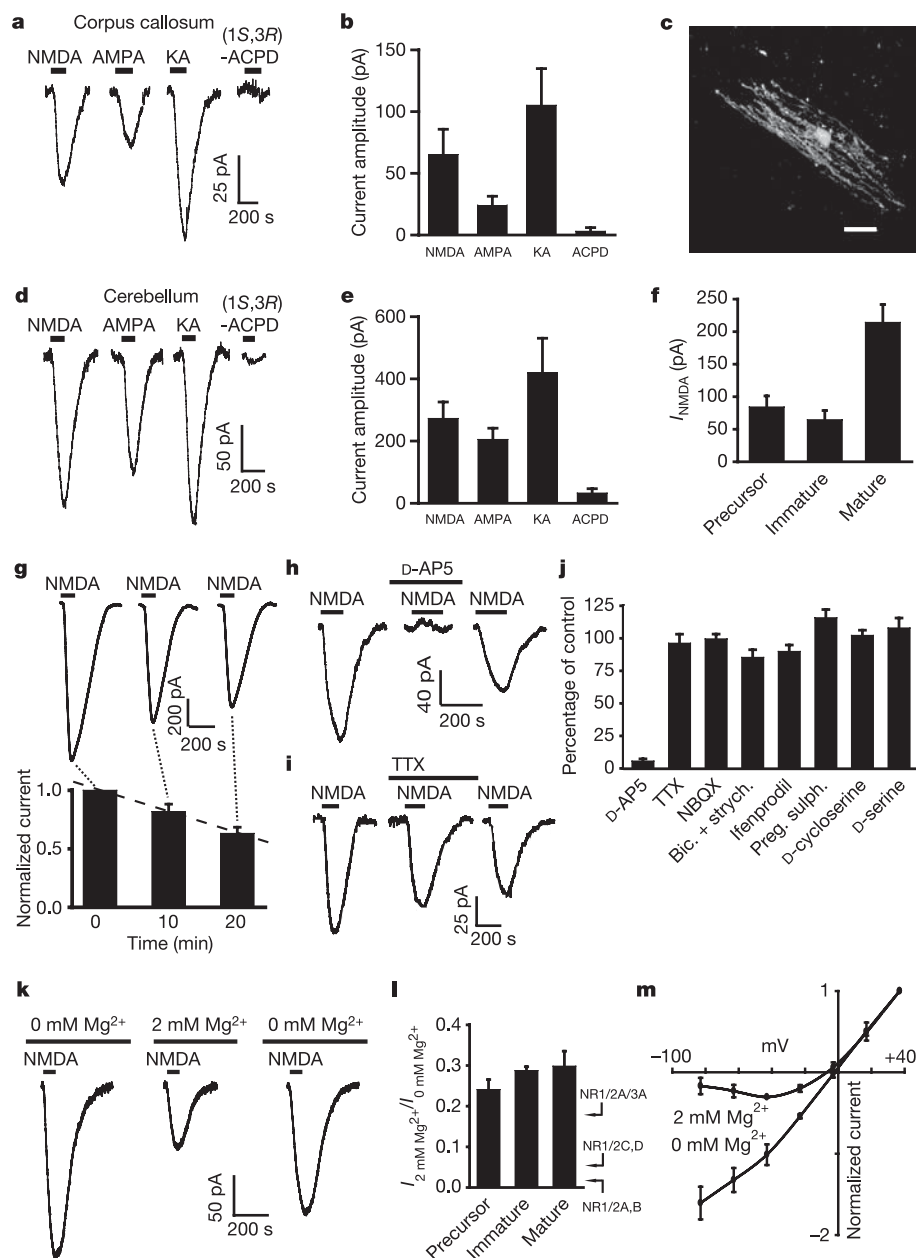


Figure 2 | Oligodendrocyte NMDA receptors show weak Mg^{2+} -block.

a, Response of a mature corpus callosum oligodendrocyte to 60 μM NMDA, 20 μM AMPA, 30 μM kainate (KA) and 100 μM (1S,3R)-ACPD (0 mM Mg^{2+}). **b**, Mean peak current in **a** for NMDA ($n = 22$ cells), AMPA ($n = 12$), KA ($n = 10$) and ACPD ($n = 5$). **c**, Mature Lucifer-filled oligodendrocyte in corpus callosum. Scale bar, 20 μm . **d**, Response of a mature cerebellar oligodendrocyte in 0 mM Mg^{2+} . **e**, Mean peak current in **d** for NMDA ($n = 26$), AMPA ($n = 23$), KA ($n = 5$) and ACPD ($n = 16$). **f**, NMDA-evoked current in cerebellar cells in Mg^{2+} -free solution ($P = 0.059$ and $P = 0.042$, comparing 79 mature with 19 immature or 26 precursor oligodendrocytes, respectively). **g**, NMDA responses run down linearly with time (dashed line) during repeated (1 per 10 min) applications. Top panel shows the response of a representative cell, bottom panel shows normalized data ($n = 10$ cells). **h**, **i**, Effect of 50 μM D-AP5 (**h**) and TTX (1 μM , **i**) on NMDA-evoked responses. **j**, Effect of 50 μM D-AP5 ($n = 16$; $P = 10^{-18}$),

1 μM TTX ($n = 5$; $P = 0.64$), 25 μM NBQX ($n = 4$; $P = 0.96$), 5 μM strychnine plus 20 μM bicuculline ($n = 16$; $P = 0.14$), 10 μM ifenprodil ($n = 3$; $P = 0.18$), 100 μM pregnenolone sulphate ($n = 4$; $P = 0.085$), 1 mM D-cycloserine ($n = 3$; $P = 0.6$) and 100 μM D-serine ($n = 5$; $P = 0.34$) on NMDA-evoked current at -63 mV. **k**, NMDA-evoked response in solution containing sequentially 0 mM, 2 mM and 0 mM Mg^{2+} . **l**, Fraction of NMDA-evoked current remaining in 2 mM Mg^{2+} , in 4 precursor, 6 immature and 9 mature cells (not significantly different, $P = 0.51$). Arrows indicate values previously obtained^{22,23} for NR1 with NR2C or NR2D, NR2A or NR2B, or NR2A and NR3A. **m**, Normalized I - V relationship for NMDA-evoked current in two precursor cells in 0 mM Mg^{2+} and three different precursor cells in 2 mM Mg^{2+} . Experiments on: corpus callosum (**a-c**), cerebellum (**d-m**); with 0 Mg^{2+} (**a-j**); all with 100 μM glycine and 5 μM strychnine (except for D-cycloserine and serine in **j**); 60 μM NMDA, 24°C , -63 mV (**a-m**).

by D-AP5 and NBQX (Fig. 4a). In some precursor oligodendrocytes (Fig. 4a, c), the inward current early in ischaemia included an increased frequency of synaptic current-like events (Fig. 1d), presumably reflecting exocytotic transmitter release²⁰ triggered by action potentials or an increase in axonal $[Ca^{2+}]_i$. The relative contribution of NMDA and AMPA receptors to the glutamate-mediated current was analysed in Mg^{2+} -free solution to detect NMDA receptor currents more accurately (in 2 mM Mg^{2+} , the NMDA component at -63 mV would be fourfold smaller (Fig. 2l), but would be increased in unclamped cells by the ischaemia-evoked current depolarizing cells and reducing NMDA receptor Mg^{2+} -block). The NMDA receptor-mediated current in precursor oligodendrocytes was on average 50% larger than that mediated by AMPA/kainate receptors (Fig. 4b–d), and in some cells comprised almost all of the current (Fig. 4c). In mature oligodendrocytes, the fraction of the inward current that was generated by glutamate was smaller than in precursor cells (Fig. 4d), suggesting that (as mature cells can generate larger glutamate- and NMDA-evoked currents; Figs 1f, 2f) less glutamate release occurs around mature cells during ischaemia. Further work is needed to establish the origin of the part of the inward current that is not generated by glutamate release.

Here we have shown that oligodendrocytes express functional NMDA receptors. The NMDA-evoked currents that we observe are unlikely to be produced by an agent released secondarily from neurons because they are unaffected by TTX or by block of AMPA/kainate, metabotropic glutamate receptors, GABA_A or glycine receptors (Figs 1i, 2j), and oligodendrocytes show no current in response to

application of noradrenaline, dopamine, histamine, serotonin, ATP, adenosine or acetylcholine (data not shown). Antibody labelling (Fig. 3) and the voltage-dependence of the current (Fig. 2m) show that oligodendrocytes themselves express NMDA receptors. Previous work that failed to detect NMDA receptor currents in oligodendrocytes was partly on cultured cells⁶, which may downregulate their NMDA receptor expression. Our data (Fig. 2a) contradict the absence of NMDA responses reported in corpus callosum oligodendrocytes⁷. As oligodendrocyte NMDA receptor currents occur in both the cerebellum and corpus callosum (Fig. 2), they may represent a general property of white matter oligodendrocytes. Our electrophysiological recordings are from precursor, immature and mature oligodendrocytes in postnatal day (P)7–14 rats, and remain to be extended to the adult; however, NMDA receptor subunits are also present in adult oligodendrocytes (Fig. 3e).

Oligodendrocyte NMDA receptors are likely to have a role in controlling oligodendrocyte development and myelination²⁶, and in damaging oligodendrocytes under pathological conditions. They show only weak block by Mg^{2+} at the cells' resting potential¹⁶ (Fig. 2l), and mediate part of the inward current generated in oligodendrocytes in response to simulation of the energy deprivation that occurs in periventricular leukomalacia, in stroke, and after ischaemia in spinal cord injury. Notably, NMDA receptors are present in the myelinating processes of oligodendrocytes, where the intracellular volume is small and receptor-mediated ion influx may produce large increases in intracellular ion concentration and osmotic water flux, which could disrupt myelination. The higher glutamate affinity of NMDA receptors relative to AMPA receptors makes them more likely to be activated in neurodegenerative disorders that involve a small but prolonged increase in extracellular glutamate concentration, as can occur in multiple sclerosis. Thus, oligodendrocyte NMDA receptors could contribute to causing the white matter damage that occurs when the extracellular glutamate concentration is increased in periventricular leukomalacia, spinal cord injury, multiple sclerosis and stroke^{1–4}. Indeed, in the optic nerve, activation of NMDA receptors on oligodendrocyte processes when

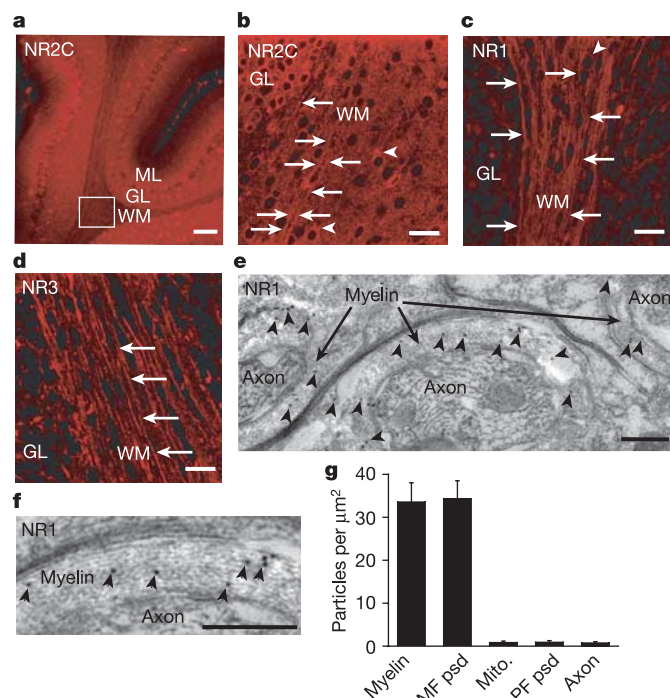


Figure 3 | Oligodendrocyte NMDA receptors. **a**, Cerebellar cortex labelled for NR2C. WM, white matter; GL, granular layer; ML, molecular layer. **b**, Enlargement of box in **a**. Arrows indicate oligodendrocyte process, arrowheads indicate oligodendrocyte soma. **c**, Oligodendrocyte process labelling with an antibody against NR1 (Chemicon; see also Supplementary Fig. 6). **d**, NR3 labelling. **e**, Immunogold (black dots indicated with arrowheads are gold particles attached to secondary antibody recognizing Wentholt NR1 antibody) labelling cerebellar myelin. **f**, Enlarged view of myelin. **g**, Gold particle density over cerebellar myelin (17 sheaths), mossy fibre synapse (MF psd, $n = 20$), parallel fibre–Purkinje synapse (PF psd, $n = 26$), mitochondria in mossy fibre terminal (mito., $n = 30$) and axon cytoplasm (axon, $n = 17$). Scale bars, 100 μm (**a**), 20 μm (**b–d**), 0.25 μm (**e**, **f**).

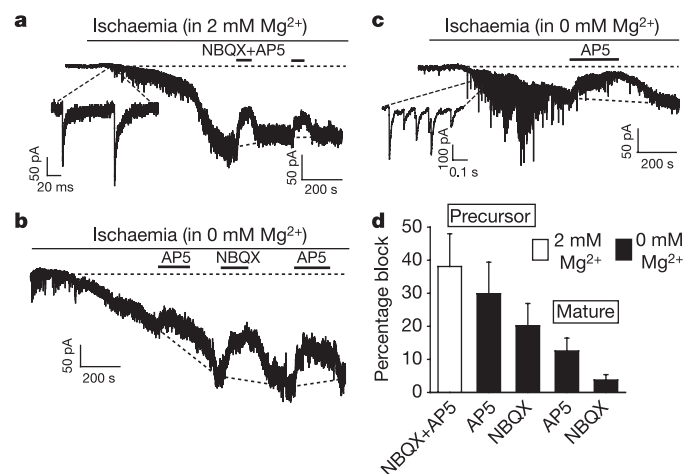


Figure 4 | Ischaemia activates NMDA receptors. **a**, Ischaemia-evoked current in precursor oligodendrocyte (in 2 mM Mg^{2+}) is blocked by 25 μM NBQX + 50 μM D-AP5. Inset shows ischaemia-induced synaptic-like currents. **b**, Precursor oligodendrocyte response to ischaemia (in 0 mM Mg^{2+}) is blocked by D-AP5, NBQX. **c**, Precursor cell response (in 0 mM Mg^{2+}), showing transient increase in synaptic currents, and most inward current blocked by D-AP5. **d**, Fractional block of inward current in precursor cells by D-AP5 + NBQX in 2 mM Mg^{2+} ($n = 7$ cells), and by D-AP5 ($n = 9$) or NBQX ($n = 7$) in 0 mM Mg^{2+} , and block of current in mature cells by D-AP5 ($n = 9$) or NBQX ($n = 10$) in 0 mM Mg^{2+} . $P = 0.098$, comparing D-AP5 in precursor and mature cells; $P = 0.019$ comparing NBQX in precursor and mature cells. **a–d**, Cerebellum, -63 mV, 33 $^{\circ}C$.

glutamate is released during ischaemia leads to the disintegration of those processes²⁷. The unusual subunit combination of these receptors, which may contain NR1, NR2C and NR3 subunits, suggests that they might be a useful therapeutic target in a variety of brain disorders.

METHODS

Brain slices. Cerebellar or forebrain slices (225- μ m thick) were prepared from P7–14 rats in solution containing 1 mM sodium kynurenate to block glutamate receptors. The cerebellum myelinates relatively late, facilitating investigation of different developmental stages. The corpus callosum is an area that becomes thinned in severe periventricular leukomalacia. Slices were superfused at $33 \pm 1^\circ\text{C}$ with bicarbonate-buffered solution (for ischaemia experiments) containing 126 mM NaCl, 24 mM NaHCO₃, 1 mM NaH₂PO₄, 2.5 mM KCl, 2 mM MgCl₂, 2.5 mM CaCl₂, 10 mM glucose, 0.1 mM glycine (which activates the NMDA receptor glycine-binding site) and 0.005 mM strychnine (which blocks glycine receptors), bubbled with 95% O₂, 5% CO₂ (pH 7.4); or at $24 \pm 1^\circ\text{C}$ with HEPES-buffered solution (for non-ischaemia experiments) containing 144 mM NaCl, 2.5 mM KCl, 2 mM MgCl₂, 10 mM HEPES, 1 mM NaH₂PO₄, 2.5 mM CaCl₂, 10 mM glucose, 0.1 mM glycine and 0.005 mM strychnine, with the pH set to 7.4 using NaOH. Omitting glycine or adding D-serine did not affect the NMDA receptor current (see text).

To simulate ischaemia, we replaced external O₂ with N₂, and external glucose with 7 mM sucrose, added 2 mM iodoacetate to block glycolysis, and added 100 μ M rotenone or 25 μ M antimycin to block oxidative phosphorylation²⁵. Without iodoacetate and rotenone/antimycin, it took ~3-fold longer for the ischaemia-evoked inward current to develop, probably because in an open chamber O₂ can diffuse to the slice, allowing glycogen metabolism in mitochondria for longer than would occur *in vivo*²⁵.

Recording and cell identification. White matter cells (avoiding cerebellar nuclei) were whole-cell clamped with pipettes containing a K⁺-based solution (for glutamate application) consisting of 130 mM KCl, 4 mM NaCl, 0.5 mM CaCl₂, 10 mM HEPES, 10 mM EGTA, 2 mM MgATP, 0.5 mM Na₂GTP, 2 mM K-Lucifer yellow, pH 7.2 (adjusted with KOH), or a Cs⁺-based solution (for NMDA application and ischaemia) consisting of 130 mM CsCl, 4 mM NaCl, 0.5 mM CaCl₂, 10 mM HEPES, 10 mM BAPTA, 2 mM MgATP, 0.5 mM Na₂GTP, 2 mM K-Lucifer yellow, pH 7.2 (adjusted with CsOH). Series resistance was 8–20 M Ω , before 60% compensation.

Cells were identified on the basis of post-recording dye-fill morphology and antibody labelling. NG2 antibody labelled 15 out of 15 tested cells with precursor morphology. (We take NG2 labelling to indicate oligodendrocyte precursors²⁰. However, some NG2 cells may be a glial class distinct from oligodendrocytes²⁸, or (in grey matter) neuronal precursors²⁹. Showing that oligodendrocytes express NMDA receptors does not depend on recording precursor cells, as NMDA receptor currents were also seen in immature and mature cells.) O4 antibody labelled 6 out of 6 cells with immature morphology. MBP antibody labelled all 17 cells with mature morphology that were tested. Antibody against the oligodendrocyte transcription factor Olig2 labelled all ten cells tested (seven precursors, one immature and two mature cells). Oligodendrocytes could be distinguished from astrocytes, which showed gap junction coupling (as revealed by Lucifer yellow spreading to other cells and GFAP labelling; 5 out of 5 cells). Electrode junction potentials were compensated. *I*–*V* relations were from responses to 200-ms voltage steps.

Immunocytochemistry. Antibody labelling and post-embedding electron microscopic immunocytochemistry are described in the Supplementary Material.

Statistics. Data are presented as mean $\pm$ s.e.m. *P* values are from Student's two-tailed *t*-tests except for multiple comparisons, which were done using one-way analysis of variance (ANOVA) and Tukey's post-hoc tests.

Received 26 July; accepted 10 October 2005.

- Volpe, J. J. Neurobiology of periventricular leukomalacia in the premature infant. *Pediatr. Res.* **50**, 553–562 (2001).
- Stys, P. K. White matter injury mechanisms. *Curr. Mol. Med.* **4**, 113–130 (2004).
- Matute, C. *et al.* The link between excitotoxic oligodendroglial death and demyelinating diseases. *Trends Neurosci.* **24**, 224–230 (2001).
- Dewar, D., Underhill, S. M. & Goldberg, M. P. Oligodendrocytes and ischemic brain injury. *J. Cereb. Blood Flow Metab.* **23**, 263–274 (2003).
- Choi, D. W. Glutamate neurotoxicity and diseases of the nervous system. *Neuron* **1**, 623–634 (1988).
- Patneau, D. K., Wright, P. W., Winters, C., Mayer, M. L. & Gallo, V. Glial cells of the oligodendrocyte lineage express both kainate- and AMPA-preferring subtypes of glutamate receptor. *Neuron* **12**, 357–371 (1994).
- Berger, T., Walz, W., Schnitzer, J. & Kettenmann, H. GABA- and glutamate-activated currents in glial cells of the mouse corpus callosum slice. *J. Neurosci. Res.* **31**, 21–27 (1992).
- Oka, A., Belliveau, M. J., Rosenberg, P. A. & Volpe, J. J. Vulnerability of oligodendroglia to glutamate: pharmacology, mechanisms, and prevention. *J. Neurosci.* **13**, 1441–1453 (1993).
- Follett, P. L., Rosenberg, P. A., Volpe, J. J. & Jensen, F. E. NBQX attenuates excitotoxic injury in developing white matter. *J. Neurosci.* **20**, 9235–9241 (2000).
- Tekkök, S. B. & Goldberg, M. P. AMPA/kainate receptor activation mediates hypoxic oligodendrocyte death and axonal injury in cerebral white matter. *J. Neurosci.* **21**, 4237–4248 (2001).
- Agrawal, S. K. & Fehlings, M. G. Role of NMDA and non-NMDA ionotropic glutamate receptors in traumatic spinal cord axonal injury. *J. Neurosci.* **17**, 1055–1063 (1997).
- Wrathall, J. R., Teng, Y. D. & Marriott, R. Delayed antagonism of AMPA/kainate receptors reduces long-term functional deficits resulting from spinal cord trauma. *Exp. Neurol.* **145**, 565–573 (1997).
- Pitt, D., Werner, P. & Raine, C. S. Glutamate excitotoxicity in a model of multiple sclerosis. *Nature Med.* **6**, 67–70 (2000).
- Smith, T., Groom, A., Zhu, B. & Turski, L. Autoimmune encephalomyelitis ameliorated by AMPA antagonists. *Nature Med.* **6**, 62–66 (2000).
- Wang, C. *et al.* Functional N-methyl-D-aspartate receptors in O-2A glial precursor cells: a critical role in regulating polysialic acid-neuronal cell adhesion molecule expression and cell migration. *J. Cell Biol.* **135**, 1565–1581 (1996).
- Ziak, D., Chvatal, A. & Sykova, E. Glutamate, kainate and NMDA-evoked membrane currents in identified glial cells in rat spinal cord slice. *Physiol. Res.* **47**, 365–375 (1998).
- Schäbitz, W.-R., Li, F. & Fisher, M. The N-methyl-D-aspartate antagonist CNS 1102 protects cerebral gray and white matter from ischemic injury following temporary focal ischemia in rats. *Stroke* **31**, 1709–1714 (2000).
- Wallström, E. *et al.* Memantine abrogates neurological deficits, but not CNS inflammation, in Lewis rat experimental autoimmune encephalomyelitis. *J. Neurol. Sci.* **137**, 89–96 (1996).
- Back, S. A. *et al.* Late oligodendrocyte progenitors coincide with the developmental window of vulnerability for human perinatal white matter injury. *J. Neurosci.* **21**, 1302–1312 (2001).
- Bergles, D. E., Roberts, J. D. B., Somogyi, P. & Jahr, C. E. Glutamatergic synapses on oligodendrocyte precursor cells in the hippocampus. *Nature* **405**, 187–191 (2000).
- Kirson, E. D., Schirra, C., Konnerth, A. & Yaari, Y. Early postnatal switch in magnesium sensitivity of NMDA receptors in rat CA1 pyramidal cells. *J. Physiol. (Lond.)* **521**, 99–111 (1999).
- Kuner, T. & Schoepfer, R. Multiple structural elements determine subunit specificity of Mg²⁺ block in NMDA receptor channels. *J. Neurosci.* **16**, 3549–3558 (1996).
- Sasaki, Y. F. *et al.* Characterization and comparison of the NR3A subunit of the NMDA receptor in recombinant systems and primary cortical neurons. *J. Neurophysiol.* **87**, 2052–2063 (2002).
- Borges, K. & Kettenmann, H. Blockade of K⁺ channels induced by AMPA/kainate receptor activation in mouse oligodendrocyte precursor cells is mediated by Na⁺ entry. *J. Neurosci. Res.* **42**, 579–593 (1995).
- Allen, N. J., Kárádóttir, R. & Attwell, D. A preferential role for glycolysis in preventing the anoxic depolarization of rat hippocampal area CA1 pyramidal cells. *J. Neurosci.* **25**, 848–859 (2005).
- Yuan, X., Eisen, A. M., McBain, C. J. & Gallo, V. A role for glutamate and its receptors in the regulation of oligodendrocyte development in cerebellar tissue slices. *Development* **125**, 2901–2914 (1998).
- Salter, M. G. & Fern, R. NMDA receptors are expressed in developing oligodendrocyte processes and mediate injury. *Nature* doi:10.1038/nature04301 (this issue).
- Berry, M., Hubbard, P. & Butt, A. M. Cytology and lineage of NG2-positive glia. *J. Neurocytol.* **31**, 457–467 (2002).
- Chittajulu, R., Aguirre, A. & Gallo, V. NG2-positive cells in the mouse white and grey matter display distinct physiological properties. *J. Physiol. (Lond.)* **561**, 109–122 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Rowitch, C. D. Stiles and J. Alberta for Olig2 antibody, F. A. Stephenson, R. J. Wenthold and O. P. Ottersen for NR1 antibody, and A. Gibb, K. Jessen, R. Mirsky, W. Richardson, D. Rossi, J. Rothman, A. Silver and J. Storm-Mathisen for advice. This work was supported by the Wellcome Trust, the European Union, the Norwegian Research Council and a Wolfson-Royal Society Award. R.K. was in the 4-year PhD Programme in Neuroscience at UCL.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A. (D.Attwell@ucl.ac.uk).

NMDA receptors are expressed in developing oligodendrocyte processes and mediate injury

Michael G. Salter¹ & Robert Fern¹

Injury to oligodendrocyte processes, the structures responsible for myelination, is implicated in many forms of brain disorder^{1–4}. Here we show NMDA (*N*-methyl-D-aspartate) receptor subunit expression on oligodendrocyte processes, and the presence of NMDA receptor subunit messenger RNA in isolated white matter. NR1, NR2A, NR2B, NR2C, NR2D and NR3A subunits showed clustered expression in cell processes, but NR3B was absent. During modelled ischaemia, NMDA receptor activation resulted in rapid Ca^{2+} -dependent detachment and disintegration of oligodendroglial processes in the white matter of mice expressing green fluorescent protein (GFP) specifically in oligodendrocytes (CNP-GFP mice). This effect occurred at mouse ages corresponding to both the initiation and the conclusion of myelination. NR1 subunits were found mainly in oligodendrocyte processes, whereas AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)/kainate receptor subunits were mainly found in the somata. Consistent with this observation, injury to the somata was prevented by blocking AMPA/kainate receptors, and preventing injury to oligodendroglial processes required the blocking of NMDA receptors. The presence of NMDA receptors in oligodendrocyte processes explains why previous studies that have focused on the somata have not detected a role for NMDA receptors in oligodendrocyte injury. These NMDA receptors bestow a high sensitivity to acute injury and represent an important new target for drug development in a variety of brain disorders.

Oligodendrocyte injury is central to the loss of function experienced in conditions ranging from cerebral palsy to spinal cord injury

and multiple sclerosis^{1–4}. Damage to oligodendrocytes is also an important secondary factor in neurological disorders such as stroke³ and Alzheimer's disease⁵. Oligodendrocytes are responsible for myelinating axons, and loss of myelination under such conditions contributes to brain dysfunction. Myelination is carried out by oligodendrocyte cell processes, and damage to these processes precedes damage to the somata in a number of disease models (for example, see refs 1, 2, 6, 7). This may indicate a heightened sensitivity to injury in these structures. In human diseases, 'dying back' of oligodendrocyte processes may be seen (for example, in multiple sclerosis^{8,9}), suggesting that this is a clinically relevant phenomenon.

Oligodendrocytes express Ca^{2+} -permeable glutamate receptors and have low resistance to oxidative stress, two factors that make them particularly susceptible to injury^{3,6,10–12}. They are thought to express mainly non-NMDA glutamate receptors, and this expression is developmentally regulated^{13,14}. The high expression of non-NMDA receptors in immature oligodendrocytes and low expression of the calcium-impermeable GluR2 subunit at the point when they initiate myelination may increase their sensitivity to an excitotoxic cascade mediated by ischaemic glutamate release and subsequent intracellular Ca^{2+} ($[\text{Ca}^{2+}]_i$) overload^{3,13,14}. This may explain the selective injury of precursor oligodendrocytes and subsequent hypomyelination in periventricular leukomalacia (PVL). PVL is the main injury associated with cerebral palsy, the most common human birth disorder. The long-term consequences of PVL can involve either focal oligodendrocyte loss (associated with early loss of cell processes

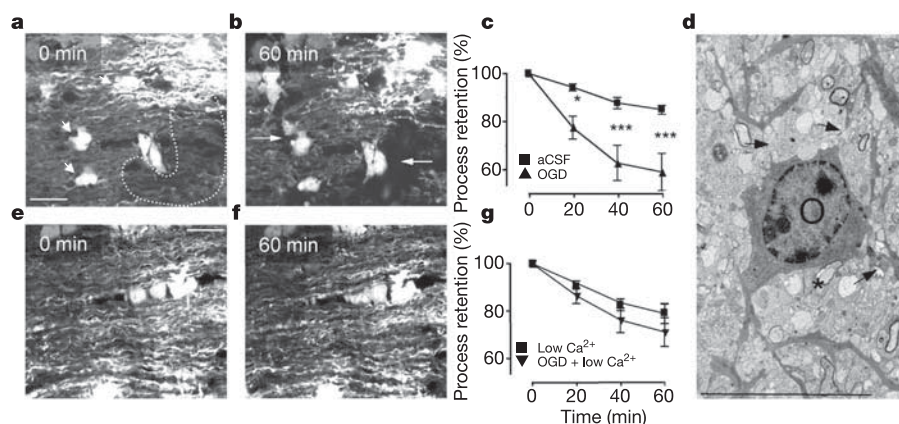


Figure 1 | Ischaemia results in rapid loss of oligodendroglial process. **a, b**, CNP-GFP oligodendrocytes before (0 min) and after 60 min OGD. Processes are lost over time (arrowheads show oligodendroglial somata, arrows show areas denuded of processes). **c**, Process retention during perfusion with artificial cerebrospinal fluid (aCSF) (squares) or during OGD (triangles). Statistical significance is shown for OGD versus aCSF

(*, $P \leq 0.005$; ***, $P \leq 0.001$), determined by ANOVA. Error bars show s.e.m. **d**, Oligodendrocyte (labelled 'O') ultrastructure after 60 min OGD. Arrows indicate detached processes, asterisk shows a myelinated axon. **e, f**, Control perfusion at 0 min and 60 min. **g**, Process retention, showing no significant difference between OGD and OGD under conditions of low Ca^{2+} ($P > 0.01$). Scale bars, 10 μm .

¹Department of Cell Physiology and Pharmacology, University of Leicester, Leicester LE1 9HN, UK.

in animal models) or diffuse disruption of myelination, associated with abnormal oligodendrocyte process morphology¹.

Live mouse optic nerve oligodendroglia expressing GFP under the control of the 2'-3'-cyclic nucleotide 3'-phosphodiesterase gene promoter (CNP) were imaged by confocal microscopy at postnatal day 10 (P10). The development of rodent white matter is delayed compared with humans, and this preparation correlates developmentally to the fetal white matter susceptible to PVL^{15,16}. The

morphology of the oligodendrocytes conformed to previous descriptions, characterized by primary processes radiating from the somata and numerous myelinating processes running parallel to axons¹⁷. Process morphology was visible in detail and was generally stable for periods of >60 min during control perfusion (Fig. 1c, e, f). This was assessed both by monitoring mean pixel intensity in large regions of interest drawn around somata-free areas (for example, dashed line in Fig. 1a), and by visual inspection using a graded scoring system (by an observer blind to the experimental conditions; see Methods).

To examine changes in oligodendrocyte morphology, the optic nerves were subjected to oxygen-glucose deprivation (OGD). During OGD, loss of processes occurred rapidly, with significant loss at 20 min and further deterioration at 40 min and 60 min (Fig. 1a–c). After 60 min of OGD, large areas denuded of GFP-stained processes were apparent (arrows in Fig. 1b). As processes deteriorated, they frequently detached from neighbouring regions of the same process and/or the somata while retaining GFP fluorescence, indicating the temporary formation of membrane-delineated structures. Formation of swellings within processes was also common. In line with previous studies, ultrastructural examination of oligodendroglial cells that survived 60 min of OGD revealed a loss of cell processes in many cases⁶ (Fig. 1d and Supplementary Fig. 1). Process detachment was confirmed by examination of serial sections. Loss of processes during ischaemia was abolished by lowering extracellular Ca^{2+} to $\sim 30 \mu\text{M}$ ($90 \mu\text{M}$ EGTA + $120 \mu\text{M}$ Ca^{2+}) (Fig. 1g) (note that low Ca^{2+} reduces fluorescence somewhat). Removing extracellular Ca^{2+} attenuated acute oligodendroglial cell death during OGD^{6,11} and had a similar protective effect on oligodendroglial somata in the optic nerve (Fig. 2i).

Using $30 \mu\text{M}$ NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulphonamide) to block Ca^{2+} influx through AMPA/kainate receptors during OGD had no effect upon the loss of processes (Fig. 2a, b, e) but prevented the loss of oligodendrocyte somata (Fig. 2i). Therefore, although OGD-induced injury to oligodendrocyte processes is Ca^{2+} -dependent, unlike in the somata the route of Ca^{2+} influx is not AMPA/kainate receptors. The widespread loss of processes resulting from OGD in the presence of NBQX was characterized by detachment of membrane-delineated processes close to the soma and by the presence of large vacuoles within processes (Fig. 2g). The presence of vacuoles in these degenerating processes might correspond to the bright GFP puncta that form during OGD, and might represent the loci where processes are detaching to form temporary membrane-delineated fragments.

The unexpected injury to processes in the presence of NBQX led us to investigate alternative potential sources of Ca^{2+} influx. When the selective NMDA receptor blocker MK801 ($10 \mu\text{M}$) was co-applied with NBQX, injury to processes was largely prevented (Fig. 2c–e) and oligodendroglia retained numerous fine processes attached to the somata (Fig. 2h). Significant protection was also achieved by MK801

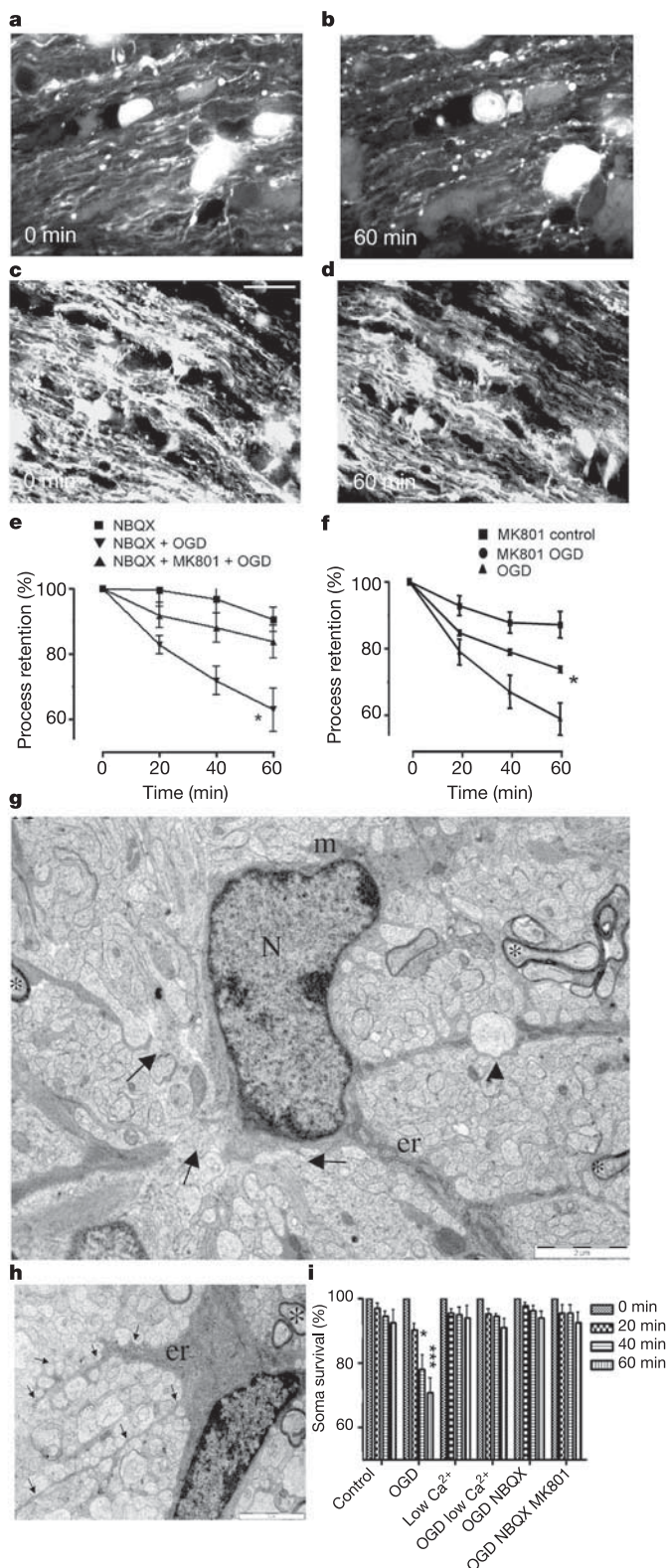


Figure 2 | NMDA receptor activation destroys oligodendrocyte processes.

a, b, Perfusion with NBQX before and after 60 min OGD. **c, d**, NBQX+MK801 before and after 60 min OGD. **e**, Process retention during treatment with NBQX (significance tested by ANOVA versus control aCSF), NBQX+OGD (tested versus OGD alone) or NBQX+MK801+OGD (tested versus NBQX+OGD). **f**, Perfusion with MK801 under normoxic conditions had no significant effect. MK801 in the absence of NBQX was significantly protective against process loss during OGD (tested versus OGD). Error bars in **e, f** indicate s.e.m.; *, $P \leq 0.005$. **g**, Oligodendrocyte exposed to OGD for 60 min in the presence of NBQX. Arrows indicate detached processes. N, nucleus; m, mitochondria; er, endoplasmic reticulum; asterisk, myelinating axon. **h**, Oligodendrocyte exposed to NBQX+MK801+OGD for 60 min. Note the intact processes radiating from the soma (arrows). **i**, Oligodendroglial soma death at four time points (0, 20, 40, 60 min) under various conditions. Error bars indicate s.e.m.; significance tested versus control; *, $P \leq 0.005$; ***, $P \leq 0.001$. Scale bars, $10 \mu\text{m}$ (**a–d**), $2 \mu\text{m}$ (**g, h**).

in the absence of NBQX, but MK801 had no effect when applied under normoxic conditions (Fig. 2f). Although ifenprodil (10 μ M), an NMDA antagonist with selectivity for the NR2B subunit, was mildly protective at $t = 20$ min when applied with NBQX

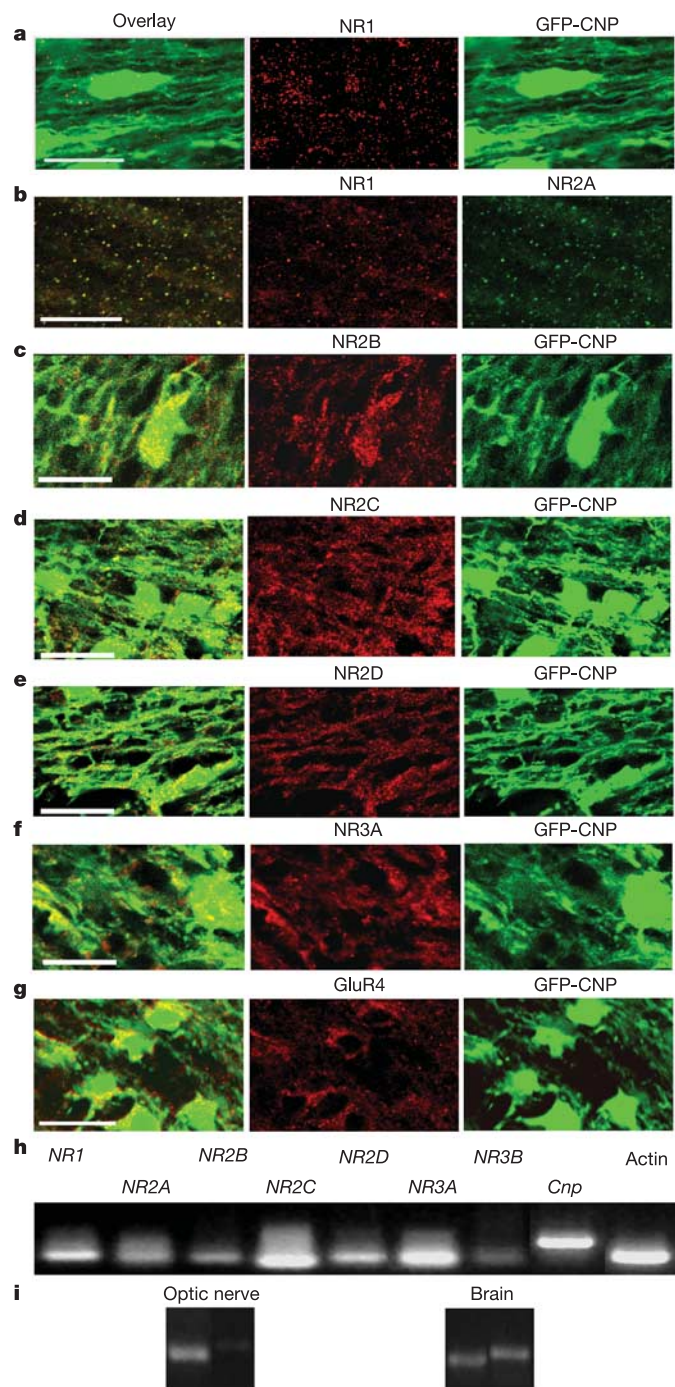


Figure 3 | Oligodendrocytes express NMDA receptors in clusters on cell processes and AMPA/kainate receptors diffusely on somata. **a**, NR1 (red) is expressed in clusters mainly in oligodendroglial (green) processes. **b**, NR1 (red) and NR2A (green) in wild-type optic nerve. **c–f**, NR2B (**c**, red), NR2C (**d**, red), NR2D (**e**, red) and NR3A (**f**, red) expression in CNP-GFP optic nerve, showing expression on oligodendrocytes (green). **g**, GluR4 (red) expression in oligodendrocyte somata (green) in CNP-GFP optic nerve. **h**, PCR products in whole optic nerve, showing the presence of *NR1*, *NR2A*, *NR2B*, *NR2C*, *NR2D* and *NR3A* NMDA receptor subunits, and the virtual absence of *NR3B*. **i**, PCR products for intron regions of *NR1* (left lane) and *Thy1* (right lane) in whole brain and optic nerve. The *NR1* intron is present in optic nerve and brain, but the *Thy1* intron is effectively present only in brain. Scale bars, 10 μ m. See Supplementary Fig. 5 for control staining.

(Supplementary Fig. 2a; $P < 0.05$), it did not protect at 60 min. Perfusion with 1 mM NMDA (+10 μ M glycine, no Mg^{2+}) under normoxic conditions did not result in significant damage to oligodendrocyte processes (Supplementary Fig. 2b), presumably reflecting the capacity of the processes to buffer Ca^{2+} influx when an adequate energy supply is available.

The above data suggest differential expression of ionotropic glutamate receptors on oligodendrocytes, with AMPA/kainate receptors expressed on somata and NMDA receptors on processes. GFP fluorescence was retained in optic nerves fixed with paraformaldehyde, and clustered expression of the NMDA receptor subunit NR1 was detected along oligodendroglial processes, using antibodies against two different regions of the NR1 subunit (Fig. 3a and Supplementary Fig. 3a). Blinded counting (by a naive observer) of NR1 clusters yielded a $57.1 \pm 4.4\%$ (mean $\pm$ s.e.m.) overlap with GFP ($n = 6$ sections). NR1 subunit expression was absent from astrocytes, as examined in P10 mice expressing GFP under the control of the GFAP promoter¹⁸ (Supplementary Fig. 4b). NR1 clusters were seen in $27.8 \pm 4.4\%$ of neurofilament-200-positive axons in wild-type FVB/N mice ($n = 4$ sections; Supplementary Fig. 4a). The occasional presence of NR1 clusters in neurofilament-200-positive axons is expected, given the close apposition of axons and oligodendrocyte processes and the resolving power of confocal imaging.

NR2A, -B, -C, -D and NR3A subunit expression was also seen in oligodendrocytes (Fig. 3b–f) but NR3B staining was absent. Double-labelling for NR1 and NR2A in wild-type optic nerve revealed a high degree of subunit co-localization on processes (Fig. 3b). The non-NMDA glutamate receptor subunit GluR4 was robustly expressed mainly in oligodendroglial somata (Fig. 3g), as were GluR2/3 subunits, but to a lesser extent (Supplementary Fig. 3b). A similar, largely somatic expression of AMPA/kainate receptors, with low or no expression on cell processes, has been noted in immature oligodendrocytes in other *in situ* preparations (for example, ref. 19).

Previous studies have failed to detect *NR1* mRNA in the rat optic nerve²⁰. Here we used fluorescence polymerase chain reaction with reverse transcription (RT-PCR), coupled to a high stringency RNA extraction protocol involving three sequential purifications steps (see Methods). We found *NR1* transcript in the optic nerve (Fig. 3h), which when quantified was at 1–2% of the abundance found in the whole brain (Supplementary Fig. 6; see ref. 21). Extraction blanks run alongside samples were negative for both *NR1* and actin, and the *NR1* product was sequenced and found to correlate to the correct portion

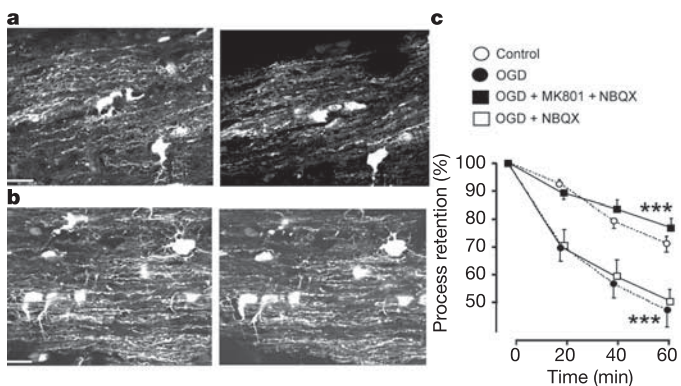


Figure 4 | Ischaemia results in rapid loss of oligodendroglial process in P25 optic nerve. **a**, Process loss in oligodendrocytes during OGD in the presence of NBQX, at 0 min (left) and 60 min (right). **b**, Process loss during OGD in the presence of NBQX+MK801, at 0 min (left) and 60 min (right). Note that in both cases the somata are retained but processes are lost when MK801 is absent. **c**, Process retention during OGD treatment (significance tested by ANOVA versus control), NBQX+OGD (tested versus OGD) or NBQX+MK801+OGD (tested versus NBQX+OGD). ***, $P \leq 0.001$. Error bars indicate s.e.m. Scale bars, 10 μ m.

of the *NR1* complementary DNA. PCR quantification revealed 2,000–3,000-fold lower abundance of the neuron-specific marker *Thy1* in the optic nerve compared with the oligodendrocyte-specific marker CNPase (this ratio was approximately 1:1 in whole brain), indicating a corresponding level of certainty that the optic nerve PCR product is glial in origin. The presence of trace *Thy1* mRNA levels within axons is consistent with previous observations showing protein synthesis within axons²².

To examine the origin of *NR1* and *Thy1* mRNA in the optic nerve, we performed RT-PCR for intron regions of *NR1* and *Thy1*, which exist only in the nucleus. This revealed robust expression of the *NR1* intron in whole brain and optic nerve, whereas expression of the *Thy1* intron was detected in whole brain but not in the optic nerve (Fig. 3i). Optic nerve *Thy1* mRNA is therefore produced in somata that are not present in the nerve (that is, retinal ganglion cells), whereas *NR1* is produced in optic nerve glial somata. PCR analysis of all known NMDA receptor subunits revealed the presence of mRNA for the subunits detected by antibody staining in oligodendrocytes (Fig. 3h). Quantification relative to actin levels suggests that *NR1*, *NR2C* and *NR3A* mRNA are the most abundant subunits in whole optic nerve; *NR2B* mRNA was present at low abundance, and *NR3B* was effectively absent.

In addition to cerebral palsy, oligodendrocyte process injury is relevant to adult diseases such as stroke, spinal cord injury and multiple sclerosis²⁴. We therefore examined process loss in P25 CNP-GFP mouse optic nerve, a stage by which effectively all precursor cells have progressed to the mature oligodendrocyte phenotype²³. Antibody staining revealed a similar pattern of *NR1*, *NR2A* and *NR2B* expression in oligodendrocytes at this stage, with abundant GluR2/3 expression mainly in somata (Supplementary Fig. 7). GFP fluorescence was relatively stable in these mature oligodendrocytes during control perfusion, but OGD evoked rapid and severe process loss (Fig. 4c). As with the P10 nerve, the loss of processes was not significantly affected by NBQX but was abolished by NBQX+MK801 (Fig. 4a–c).

A companion paper to this one (ref. 24) describes NMDA receptor currents in oligodendrocytes in several brain regions and at various developmental stages. These NMDA receptor-mediated currents show a low degree of voltage-dependent Mg^{2+} block, and immunostaining results indicate that *NR1*, *NR2C* and *NR3* are the major NMDA receptor subunits, but *NR2A* and *NR2B* are also present. Our PCR and immunostaining results show similar patterns of expression of the main subunits, and the presence of robust *NR2C* and *NR3A* subunit expression indicates a similar, low Mg^{2+} block. Several reports suggest low Ca^{2+} permeability in NMDA receptors that incorporate the *NR1A*, *NR2A*/*NR2B* and *NR3A* subunits^{25,26}. However, no information is available regarding the Ca^{2+} permeability of receptors that include the *NR2C* subunit, which may not share this feature when incorporated with *NR3A*. Furthermore, the dimensions of oligodendrocyte processes are small, and even NMDA receptors with low Ca^{2+} permeability may raise intracellular Ca^{2+} to toxic levels within such a confined space. Our data clearly show the Ca^{2+} -dependence of NMDA receptor-mediated process loss, and indicate that sufficient Ca^{2+} influx occurs through the NMDA receptors on processes to result in injury.

There is evidence for the presence of NMDA receptors on mature astrocytes, Muller cells and Bergman glia¹³, and some evidence for their presence on oligodendrocytes in other preparations^{27,28}. Activation of the AMPA/kainate receptors expressed by developing oligodendroglia can influence gene transcription and cell proliferation, survival and fate¹³. The functional significance of NMDA receptor expression in immature oligodendroglial processes is unclear and might involve axon–glial signalling during myelogenesis. Myelin initiation begins with the extension of multiple processes from the somata that make contact with axons¹⁷. At this point, oligodendroglial processes will either proceed with myelination or retract from the axon. How this decision is controlled is

unclear, but Ca^{2+} influx through activated NMDA receptors would affect cytoskeletal elements within the processes and could determine stabilization/retraction. Regardless of the function of NMDA receptors on developing oligodendrocyte processes, their pathophysiological relevance is high, as they confer a sensitivity to injury that is likely to have significance for a variety of neurological diseases. It is encouraging that our results show that NMDA receptor blockade alone can be sufficient to protect against injury. The unusual subunit composition of the receptors (which include mainly *NR2C* in addition to *NR3A* subunits) also raises the prospect of developing targeted interventions with fewer side effects than those experienced with non-selective NMDA antagonists.

METHODS

Animals and tissue analysis. Transgenic mice (FVB/N) carrying the enhanced GFP sequence (EGFP) under the control of mouse CNP promoters 1 and 2 (ref. 29) were donated by the laboratory of V. Gallo. Mice were backcrossed to wild-type FVB/N females to the third generation to reduce GFP expression levels. Animals were maintained in accordance with UK Home Office regulations. Optic nerves were dissected at P7–13, sealed in an atmospheric perfusion chamber at 37 °C and imaged using an Olympus IX70 confocal microscope. Optic nerves were perfused and exposed to ischaemic conditions using established protocols (see ref. 6).

Loss of GFP-filled processes was assessed as a change in mean pixel intensity within about three large representative (in terms of process density and brightness), soma-free regions of interest (ROI, see Fig. 1a) per flattened image stack (200 × 200 × 15 μm thick). This will underestimate fluorescence changes in processes owing to the incorporation of process-free areas. ROI intensity changes were analysed by analysis of variance (ANOVA), with all experiments within a group compared at each time point (Tukey's post-hoc test). For all figures, one asterisk, $P \leq 0.005$; three asterisks, $P \leq 0.001$. There was no correlation between the initial GFP brightness and the changes seen during OGD or during drug treatments. In addition, image stacks were scored by eye (blind to treatment) for process damage, using the following scoring system: 0, no change apparent; 1, detectable loss of processes; 2, clear loss of processes; 3, severe loss of processes; 4, a few processes retained; 5, only a few GFP puncta retained; 6, complete loss of processes. Scored analyses were tested using Kruskal–Wallis tests with Tukey's post-hoc tests, and in all cases confirmed the changes detected by pixel intensity analysis (data not shown). Imaging parameters and laser settings were unchanged throughout each experiment.

Immunohistochemistry. For immunohistochemistry, optic nerves were dissected in 0.1 M PBS and fixed in 4% paraformaldehyde for 30 min. Optic nerves were then incubated in 0.1 M PBS plus 20% sucrose (w/v) for 5 min before freeze-sectioning and subsequent incubation in 0.1 M PBS, 10% of an appropriate fetal serum, 0.5% Triton X-100 and primary antibody overnight at 4 °C. Separate polyclonal antibodies raised against the *NR1* subunit amino terminus (rabbit, Upstate) or carboxy terminus (goat, Santa Cruz) were used at 1:200. Rabbit polyclonal antibodies raised against GluR4 and GluR2/3 (Upstate) were used at 1:200. Goat polyclonal antibodies raised against *NR2A*, *NR2B*, *NR2C* and *NR2D* (Upstate) were used at between 1:100 and 1:500. Rabbit polyclonal anti-*NR3A* and anti-*NR3B* antibodies (Santa Cruz) were used at 1:100. Monoclonal anti-neurofilament-200 (NF-200) antibody (Alomone) was used at 1:100. Appropriate Alexa-conjugated secondary antibodies (Cambridge Bioscience) were used at 1:1,000 and were applied for 2 h after washing.

RT-PCR. Tissue was ground up in liquid nitrogen and subjected to a triazol extraction. RNA was then column-purified (RNeasy lipid RNA mini kit, Qiagen) and the resulting RNA treated to remove DNA contamination (DNasefree, Ambion). The RNA was purified by extraction with phenol (Sigma) pH 4.2, and resuspended to a concentration of 1 $\mu g \mu l^{-1}$. RT-PCR was performed on 1–3 μg of the RNA samples using Omniscript (Qiagen) according to the manufacturer's instructions (all quantities were doubled). The reverse-transcribed mix (2 μl) was then used for fluorescence PCR in the SYBR green system using Quantitect SYBR green PCR master mix (Qiagen) with a 40 μl reaction volume. See Supplementary Fig. 8 for primer sequences. All primers were tested on whole brain as a positive control. Primers were optimized for >90% efficiency (95 °C for 45 s, 57 °C for 30 s, 72 °C for 45 s, for 40 cycles with a dissociation curve at the end).

Electron microscopy. Optic nerves were subjected to 60 min of OGD before washing in Sorenson's buffer and fixation in 3% glutaraldehyde/Sorenson's. Nerves were post-fixed with 2% osmium tetroxide and dehydrated before infiltration in epoxy. Sections were counterstained with uranyl acetate and lead citrate, and examined using a Jeol 100CX electron microscope. Electron

micrographs were collected by R.F., who was blind to the experimental procedure used to produce each sample.

Received 27 July; accepted 10 October 2005.

- Back, S. A. *et al.* Selective vulnerability of late oligodendrocyte progenitors to hypoxia-ischemia. *J. Neurosci.* **22**, 455–463 (2002).
- Grossman, S. D., Rosenberg, L. J. & Wrathall, J. R. Temporal–spatial pattern of acute neuronal and glial loss after spinal cord contusion. *Exp. Neurol.* **168**, 273–282 (2001).
- Dewar, D., Underhill, S. M. & Goldberg, M. P. Oligodendrocytes and ischemic brain injury. *J. Cereb. Blood Flow Metab.* **23**, 263–274 (2003).
- Pitt, D., Werner, P. & Raine, C. S. Glutamate excitotoxicity in a model of multiple sclerosis. *Nature Med.* **6**, 67–70 (2000).
- Brun, A. & Englund, E. A white matter disorder in dementia of the Alzheimer type: a pathoanatomical study. *Ann. Neurol.* **19**, 253–262 (1986).
- Wilke, S., Salter, M., Thomas, R., Allcock, N. & Fern, R. Mechanism of acute ischemic injury of oligodendroglia in early myelinating white matter: the importance of astrocyte injury and glutamate release. *J. Neuropathol. Exp. Neurol.* **63**, 872–881 (2004).
- Ludwin, S. K. Pathology of demyelination and remyelination. *Adv. Neurol.* **31**, 123–168 (1981).
- Rodriguez, M., Scheithauer, B. W., Forbes, G. & Kelly, P. J. Oligodendrocyte injury is an early event in lesions of multiple sclerosis. *Mayo Clin. Proc.* **68**, 627–636 (1993).
- Wolswijk, G. Oligodendrocyte survival, loss and birth in lesions of chronic-stage multiple sclerosis. *Brain* **123**, 105–115 (2000).
- Back, S. A., Gan, X., Li, Y., Rosenberg, P. A. & Volpe, J. J. Maturation-dependent vulnerability of oligodendrocytes to oxidative stress-induced death caused by glutathione depletion. *J. Neurosci.* **18**, 6241–6253 (1998).
- Fern, R. & Moller, T. Rapid ischemic cell death in immature oligodendrocytes: a fatal glutamate release feedback loop. *J. Neurosci.* **20**, 34–42 (2000).
- Follett, P. L., Rosenberg, P. A., Volpe, J. J. & Jensen, F. E. NBQX attenuates excitotoxic injury in developing white matter. *J. Neurosci.* **20**, 9235–9241 (2000).
- Gallo, V. & Ghiani, C. A. Glutamate receptors in glia: new cells, new inputs and new functions. *Trends Pharmacol. Sci.* **21**, 252–258 (2000).
- Itoh, T. *et al.* AMPA glutamate receptor-mediated calcium signalling is transiently enhanced during development of oligodendrocytes. *J. Neurochem.* **81**, 390–402 (2002).
- Small, R. K., Riddle, P. & Noble, M. Evidence for migration of oligodendrocyte-type-2 astrocyte progenitor cells into the developing rat optic nerve. *Nature* **328**, 155–157 (1987).
- Craig, A. *et al.* Quantitative analysis of perinatal rodent oligodendrocyte lineage progression and its correlation with human. *Exp. Neurol.* **181**, 231–240 (2003).
- Butt, A. M. & Ransom, B. R. Morphology of astrocytes and oligodendrocytes during development in the intact rat optic nerve. *J. Comp. Neurol.* **338**, 141–158 (1993).
- Zhuo, L. *et al.* Live astrocytes visualized by green fluorescent protein in transgenic mice. *Dev. Biol.* **187**, 36–42 (1997).
- McDonald, J. W., Althomsons, S. P., Hyrc, K. L., Choi, D. W. & Goldberg, M. P. Oligodendrocytes from forebrain are highly vulnerable to AMPA/kainate receptor-mediated excitotoxicity. *Nature Med.* **4**, 291–297 (1998).
- Matute, C., Sanchez-Gomez, M. V., Martinez-Millan, L. & Miledi, R. Glutamate receptor-mediated toxicity in optic nerve oligodendrocytes. *Proc. Natl Acad. Sci. USA* **94**, 8830–8835 (1997).
- Salter, M. G., Franklin, K. A. & Whitelam, G. C. Gating of the rapid shade-avoidance response by the circadian clock in plants. *Nature* **426**, 680–683 (2003).
- Alvarez, J., Giuditta, A. & Koenig, E. Protein synthesis in axons and terminals: significance for maintenance, plasticity and regulation of phenotype. With a critique of slow transport theory. *Prog. Neurobiol.* **62**, 1–62 (2000).
- Barres, B. A. *et al.* Cell death and control of cell survival in the oligodendrocyte lineage. *Cell* **70**, 31–46 (1992).
- Kárádóttir, R., Cavelier, P., Bergersen, L. H. & Attwell, D. NMDA receptors are expressed in oligodendrocytes and activated in ischaemia. *Nature* doi:10.1038/nature04302 (this issue).
- Matsuda, K., Kamiya, Y., Matsuda, S. & Yuzaki, M. Cloning and characterization of a novel NMDA receptor subunit NR3B: a dominant subunit that reduces calcium permeability. *Brain Res. Mol. Brain Res.* **100**, 43–52 (2002).
- Sasaki, Y. F. *et al.* Characterization and comparison of the NR3A subunit of the NMDA receptor in recombinant systems and primary cortical neurons. *J. Neurophysiol.* **87**, 2052–2063 (2002).
- Wang, C. *et al.* Functional N-methyl-D-aspartate receptors in O-2A glial precursor cells: a critical role in regulating polysialic acid–neural cell adhesion molecule expression and cell migration. *J. Cell Biol.* **135**, 1565–1581 (1996).
- Ziak, D., Chvatal, A. & Sykova, E. Glutamate-, kainate- and NMDA-evoked membrane currents in identified glial cells in rat spinal cord slice. *Physiol. Res.* **47**, 365–375 (1998).
- Yuan, X. *et al.* Expression of the green fluorescent protein in the oligodendrocyte lineage: a transgenic mouse for developmental and physiological studies. *J. Neurosci. Res.* **70**, 529–545 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We wish to thank V. Gallo for the gift of CNP-GFP mice, I. Eperon for discussion, J. Alix for immunostaining advice, and N. Alcock for technical assistance with electron microscopy. This work was supported by a grant from the National Institutes of Neurological Disorders and Stroke to R.F.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.F. (rf34@le.ac.uk).

LETTERS

WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators

Andrea Leibfried¹, Jennifer P. C. To², Wolfgang Busch^{1*}, Sandra Stehling^{1*}, Andreas Kehle¹, Monika Demar¹, Joseph J. Kieber² & Jan U. Lohmann¹

Plants continuously maintain pools of totipotent stem cells in their apical meristems from which elaborate root and shoot systems are produced. In *Arabidopsis thaliana*, stem cell fate in the shoot apical meristem is controlled by a regulatory network that includes the CLAVATA (CLV) ligand–receptor system and the homeodomain protein WUSCHEL (WUS)^{1,2}. Phytohormones such as auxin and cytokinin are also important for meristem regulation³. Here we show a mechanistic link between the CLV/WUS network and hormonal control. WUS, a positive regulator of stem cells, directly represses the transcription of several two-component ARABIDOPSIS RESPONSE REGULATOR genes (ARR5, ARR6, ARR7 and ARR15), which act in the negative-feedback loop of cytokinin signalling^{4,5}. These data indicate that ARR genes might negatively influence meristem size and that their repression by WUS might be necessary for proper meristem function. Consistent with this hypothesis is our observation that a mutant ARR7 allele, which mimics the active, phosphorylated form, causes the formation of aberrant shoot apical meristems. Conversely, a loss-of-function mutation in a maize ARR homologue was recently shown to cause enlarged meristems⁶.

Genetic analyses have led to the discovery of several essential regulators of stem cell fate in the shoot apical meristem of the reference

plant *Arabidopsis thaliana*. Among them, the homeodomain transcription factors WUSCHEL (WUS) and SHOOTMERISTEM-LESS (STM) have positive functions^{7,8}, whereas the CLAVATA (CLV) genes negatively influence meristem size^{9–11}. WUS is expressed in the organizing centre and induces stem cell fate in the overlying cells¹² that in turn express CLV3, a small secreted peptide^{13,14} that is thought to act as ligand for the CLV1–CLV2 heteromeric receptor complex^{15,16}. Activation of the CLV1–CLV2 receptor leads to the suppression of WUS expression, creating a negative feedback loop that controls the size of the stem cell pool^{1,2}.

Despite the central role of the WUS transcription factor in the initiation and maintenance of stem cell fate, only a single direct target, the floral homeotic gene AGAMOUS (AG), which represses the maintenance of stem cells in the flower, has been described¹⁷. To identify target genes of WUS and other meristem regulators, we performed a comparative microarray screen using plants with ethanol-inducible overexpression alleles¹⁸ of WUS as well as STM and LEAFY (LFY), a floral regulator that interacts with WUS^{17,19}. After 12 h of treatment with ethanol we harvested the shoot apex and surrounding tissue (Fig. 1a) and subjected it to expression profiling with Affymetrix Ath1 arrays. A combination of per-gene and common variance²⁰ filtering was used to identify 148 genes responsive to

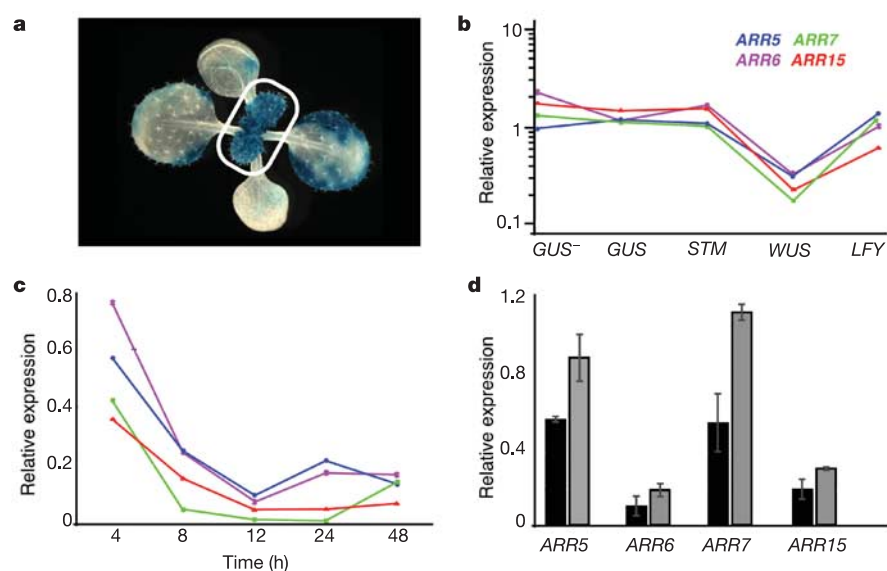


Figure 1 | Expression profiles of ARR5, ARR6, ARR7 and ARR15. **a**, A 12-day-old seedling showing ectopic AG::GUS reporter gene activation in response to WUS induction. Tissue used for expression profiling is indicated. **b**, Expression of ARR5 (blue), ARR6 (purple), ARR7 (green) and ARR15 (red) is specifically repressed by WUS as detected by microarrays. **c**, Real-time qRT-PCR confirms rapid repression of ARR genes by WUS. Relative expression is normalized to induced *Alca::GUS* controls. Line colours are as in **b**. **d**, ARR expression in response to downregulation of WUS by induction of *Alca::CLV3* (grey bars). Black bars, *Alca::GUS*. Relative expression measured by real-time qRT-PCR is normalized to *TUBULIN*. Error bars indicate s.e.m.

¹Max Planck Institute for Developmental Biology, AG Lohmann, D-72076 Tübingen, Germany. ²Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

*These authors contributed equally to this work.

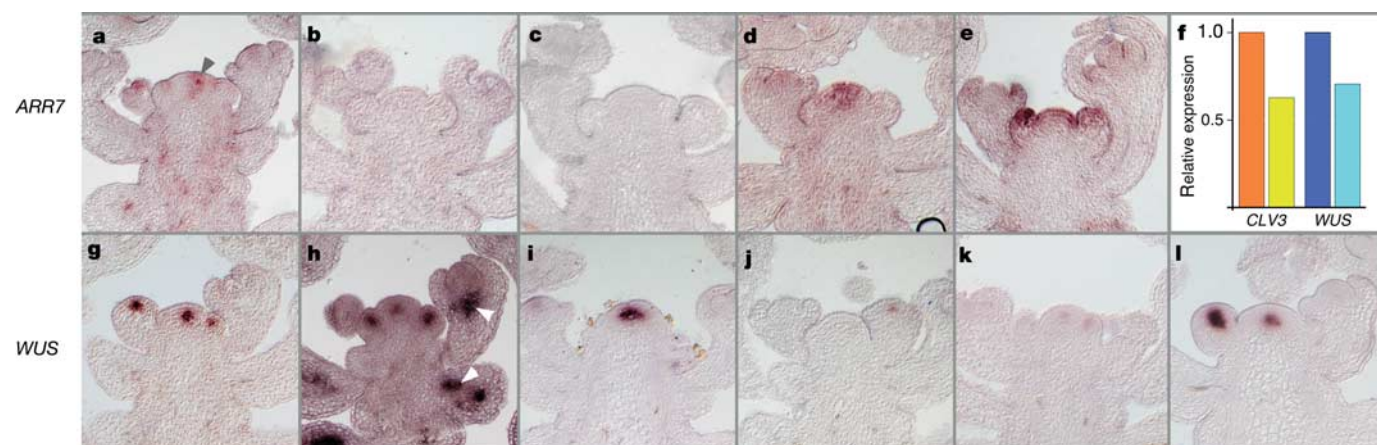


Figure 2 | Expression patterns of *ARR7* and *WUS* in response to meristematic signals. Upper panels show *in situ* hybridizations of *ARR7*; lower panels show *WUS*. **a, g**, Wild type. **b, h**, Induced *35S::AlcR AlcA::WUS*. *WUS* is moderately expressed in all cells with hot spots in more mature tissue (arrowheads). **c, i**, *clv3-7* mutant. **d, j**, Induced *35S::AlcR AlcA::CLV3*.

e, 6-Benzylaminopurine-treated wild type. **f**, *CLV3* and *WUS* expression in inflorescence apices of wild-type and *arr3 arr4 arr5 arr6 arr7 arr8 arr9* septuple mutants as measured by qRT-PCR. Dark colours represent wild-type, light colours indicate mutant. **k**, *35S::ARR7*. **l**, *arr3 arr4 arr5 arr6 arr7 arr8 arr9* septuple mutant.

WUS but not to *STM* or *LFY* induction (Supplementary Table 1). Of these 148 genes, 44 were repressed, including *ARR5*, *ARR6*, *ARR7* and *ARR15*, which belong to the 10-member type-A *ARABIDOPSIS RESPONSE REGULATOR* gene family²¹ (Fig. 1b). Type-A *ARR* proteins contain a phosphate-accepting receiver domain similar to bacterial two-component response regulators, but in contrast to type-B *ARR* proteins they lack a DNA-binding motif in their output domain²². Their expression is rapidly induced by cytokinin²¹, which has been shown to be a potent inducer of cell proliferation when applied exogenously together with auxin and to induce shoot development when acting alone²³. Type-A *ARR* proteins have been implicated in the negative feedback regulation of cytokinin signalling on the basis of the observation of decreased hormone sensitivity in plants overexpressing type-A *ARR* genes^{4,24}. Furthermore, in *Arabidopsis*, type-A *arr* multiple mutants have increased cytokinin sensitivity. However, even in sextuple type-A *arr* mutants (*arr3 arr4 arr5 arr6 arr8 arr9*) morphological changes are minimal, indicating strong redundancy within the gene family⁵. *ARR5* and *ARR6*, as well as *ARR7* and *ARR15*, constitute closely related pairs within the gene family²¹, and inspection of the AtGenExpress expression atlas²⁵ revealed co-expression of each pair, marked by widespread transcription with highest levels in meristematic tissue for *ARR7* and *ARR15*, and in roots for *ARR5* and *ARR6*.

By using quantitative real-time reverse transcriptase-mediated polymerase chain reaction (qRT-PCR), we found that 4 h after *WUS* induction by ethanol, RNA levels of *ARR5*, *ARR6*, *ARR7* and *ARR15* were already decreased, and after 12 h they reached a minimum at about 10% of control levels. Expression levels remained low for at least 48 h after treatment with ethanol (Fig. 1c). To test whether *WUS* is not only sufficient but also necessary for the repression of *ARR5*, *ARR6*, *ARR7* and *ARR15* in wild-type meristems, we used inducible *CLV3* to transiently repress *WUS*, because the morphology of *wus* mutants deviates strongly from the wild type even at very early stages of development⁷. Besides a strong reduction of *WUS* expression, we observed by qRT-PCR a moderate increase in expression of the *ARR* genes after 24 h of *CLV3* induction (Fig. 1d), which is consistent with the idea that *ARR* expression extended into the small *WUS* domain in these plants.

In situ hybridization on sections of inflorescence meristems demonstrated that *ARR7* RNA accumulates in a subdomain of the meristem consistent with a potential function in this tissue (Fig. 2a, and Supplementary Fig. 1). Reporter gene analysis confirmed this pattern and showed in addition that *ARR5*, *ARR6* and *ARR15* promoters are also active in the meristem (Supplementary Fig. 2).

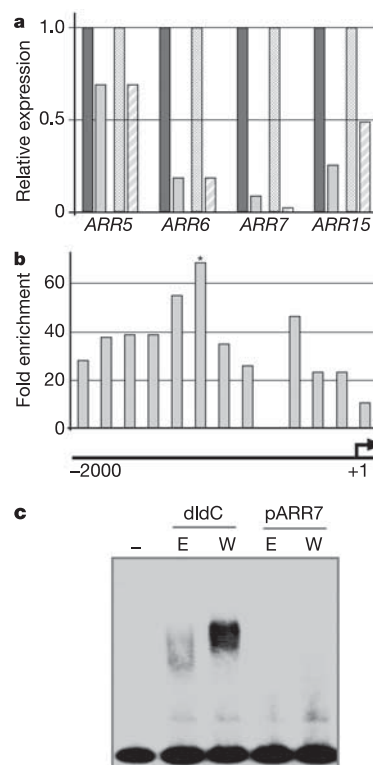


Figure 3 | Direct interaction of *WUS* with regulatory sequences of *ARR7*.

a, Real-time qRT-PCR on *35S::WUS:GR* plants. Dark grey bars represent mock treatment, light grey bars induction with dexamethasone, cross-hatched bars mock treatment in the presence of cycloheximide, and hatched bars induction with dexamethasone in the presence of cycloheximide. Expression values are normalized to the respective mock treatment controls (see Supplementary Fig. 3a for alternative normalization). **b**, Detection of *ARR7* regulatory sequences by real-time qRT-PCR after ChIP with anti-*WUS* antiserum (see Supplementary Fig. 3b). Enrichment of overlapping genomic fragments upstream of the *ARR7* start codon is shown after normalization to unrelated control sequences (see also Supplementary Fig. 3c for alternative normalization). ChIP was performed on induced *35S::WUS:GR* tissue. Asterisk, promoter fragment used for gel shifts. **c**, EMSA using *ARR7* promoter sequences identified in **b**; –, free probe; E, control protein extract from yeast expressing *LEAFY*; W, protein extract from yeast expressing *WUSCHEL*; dIdC, poly(dIdC) used as unspecific competitor; pARR7, unlabelled probe used as specific competitor.

In plants with an inducible *WUS* transgene (Fig. 2h), *ARR7* RNA could no longer be detected 24 h after *WUS* induction (Fig. 2b), which is similar to the situation in *clv3* mutants (Fig. 2c), in which *WUS* expression is expanded (Fig. 2g, i). Conversely, after suppression of *WUS* by *CLV3* induction (Fig. 2j), we observed an expansion of the *ARR7* expression domain (Fig. 2d), confirming the qRT-PCR results. Activation of *ARR7* in cells outside the *WUS* domain might indicate a more direct effect of *CLV3* on *ARR7* expression in parallel to its *WUS* dependent activity. Similarly to what has been observed for the maize homologue *ABPH1* (ref. 6), a 30-min treatment with the synthetic cytokinin 6-benzylaminopurine caused an expansion of *ARR7* expression in the wild type (Fig. 2e).

An additional level of regulation is provided by negative feedback of *ARR7* on *WUS*, because plants that overexpress *ARR7* from the constitutive 35S promoter have lower *WUS* RNA levels (Fig. 2k). However, residual *WUS* activity in 35S::*ARR7* plants is sufficient for correct function of the meristem, because 35S::*ARR7* plants have no obvious defects in the shoot apical meristem, similar to induced *AlcA::CLV3* plants, which show a *wus* mutant phenotype only in flowers (data not shown).

Having established a regulatory interaction between *WUS* and *ARR7*, we next asked whether this interaction is direct. To this end, we first made use of an inducible form of *WUS* by means of a translational fusion to the ligand-binding domain of the rat glucocorticoid receptor (*WUS:GR*). Application of a steroid such as dexamethasone causes translocation of the fusion protein from the cytoplasm to the nucleus, allowing activation or repression of direct targets in the absence of protein synthesis^{26,27}. After treatment of 35S::*WUS:GR* plants with dexamethasone for 4 h, we observed robust repression of *ARR5*, *ARR6*, *ARR7* and *ARR15*. Repression of the *ARR* genes also occurred in the presence of the protein synthesis inhibitor cycloheximide (Fig. 3a, and Supplementary Fig. 3a), which is compatible with a direct interaction of *WUS* with the regulatory elements of the *ARR* genes. We then confirmed *in vivo* binding of *WUS* to *ARR7* promoter sequences by chromatin immunoprecipitation (ChIP) with a polyclonal anti-*WUS* antiserum (Fig. 3b, and Supplementary Fig. 3b, c). We observed a twofold enrichment of *ARR7* promoter DNA in wild-type inflorescences in comparison

with leaves, in which *WUS* is not expressed, whereas in *WUS*-overexpressing tissue *ARR7* promoter DNA was enriched 68-fold. The ChIP results indicated binding of *WUS* to sequences located about 1,000 base pairs upstream of the start codon of *ARR7* in a region harbouring multiple TAAT elements, which have been shown to be the core binding sites for *WUS* (Fig. 3b)¹⁷. Subsequently, we were able to confirm sequence-specific binding of *WUS* protein to this promoter element by electrophoretic mobility-shift assays (EMSAs) (Fig. 3c).

It has recently been shown that maize mutants defective for *ABPH1*, a type-A *ARR* homologue, have defects in phyllotaxis and meristem size regulation⁶. In contrast, neither *Arabidopsis* plants lacking individual type-A *ARR* genes nor plants overexpressing *ARR5*, *ARR6*, *ARR7* or *ARR15* have obvious phenotypes (data not shown, and refs. 4, 5). We therefore constructed *arr7 arr15* double mutants, because they are closely related and both are expressed in meristematic tissue. However, the double mutant combination caused female gametophytic lethality, precluding analysis of the progeny. To reduce redundancy outside the *ARR7/ARR15* pair, we then extended our analysis to *arr3 arr4 arr5 arr6 arr7 arr8 arr9* septuple mutants. These plants were viable, although they had defects in phyllotaxis and organ initiation (Fig. 4a, b), indicating that the redundant function of *ARR7* and *ARR15* might be sufficient for meristem maintenance. *WUS* expression in the inflorescence meristem of septuple mutants was decreased (Fig. 2f, l), indicating that, in addition to the negative regulatory activity of *ARR7* on *WUS*, there might be positive effects on *WUS* expression by other type-A *ARR* genes.

As an alternative to exploring *ARR7* function, we constructed alleles that either mimic the active, phosphorylated state or the inactive non-phosphorylated state of *ARR7* by mutating aspartate 85 to glutamate or asparagine, respectively²⁸. Whereas ubiquitous overexpression of the dominant-negative form (Asp 85 → Asn) did not cause any morphological defects, the constitutively active form (Asp 85 → Glu) had severe effects on the function of the shoot apical meristem. In some of the transgenic seedlings meristems were arrested for several days after expansion of the cotyledons, resulting in an almost complete block of organ formation, very similar to that

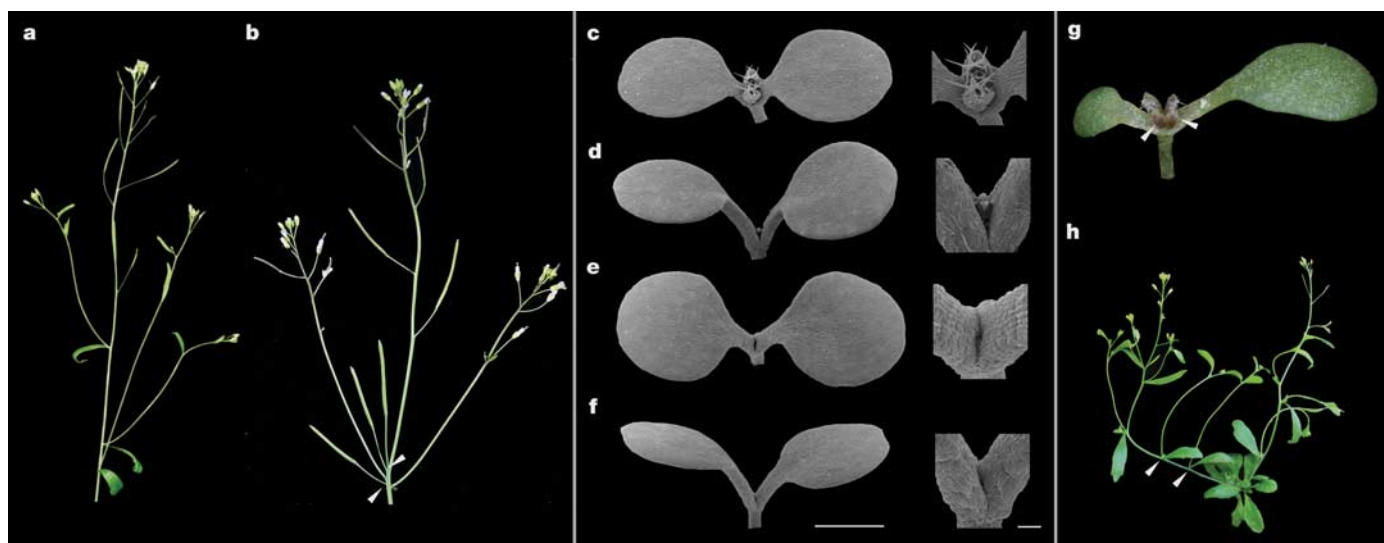


Figure 4 | Phenotypes of type-A *ARR* mutant plants. **a**, Wild type. **b**, *arr3 arr4 arr5 arr6 arr7 arr8 arr9* septuple mutant; note irregular organ positioning indicated by arrowheads. **c–f**, Activity of the shoot apical meristem is arrested in 35S::*ARR7* (Asp 85 → Glu) plants similar to *wus* mutants 5 days after sowing. Scale bars, 1 mm for seedlings and 100 μm for meristem insets unless otherwise noted. **c**, 35S::*ARR7* (Asp 85 → Glu) plant with wild-type morphology. Scale bar, 200 μm. **d**, 35S::*ARR7* (Asp 85 → Glu)

plant with intermediate phenotype. **e**, 35S::*ARR7* (Asp 85 → Glu) plant with strong phenotype. **f**, *wus* mutant seedling. **g**, 35S::*ARR7* (Asp 85 → Glu) seedling shortly after recovery of meristematic activity. Arrowheads indicate duplicated meristems. **h**, Phenotype of an adult 35S::*ARR7* (Asp 85 → Glu) plant after recovery. Note duplicated rosettes. Arrowheads indicate irregular side-shoot positions.

observed in *wus* mutants (Fig. 4c–f). Subsequently, shoot apical meristems recovered proliferative activity, but often split into two or three independent meristems (Fig. 4g), giving rise to multiple primary shoots. Similarly to the *abph1* mutant of maize⁶, these shoots had defects in phyllotaxis (Fig. 4h) and flower formation; in addition they did not produce seeds.

Our results show that direct interaction between the CLV/WUS network and the cytokinin signalling circuitry is required for proper meristem function. Together with the recently uncovered role of the type-A response regulator *ABPH1* in maize⁶, our findings are a first step towards understanding how global hormonal signals are integrated with local transcriptional inputs in the regulation of cell behaviour at the shoot apical meristem.

METHODS

Plant material and treatments. Plants were of Columbia background and grown at 23°C in continuous light. Inductions with ethanol were performed at 20°C by watering with 1% ethanol. For inductions with dexamethasone, tissue was incubated in 15 µM dexamethasone and 0.015% Silwet L-77. Cycloheximide was used at 10 µM. For 6-benzylaminopurine treatments, tissue was incubated in 1 µM 6-benzylaminopurine and 0.1% DMSO. The Columbia *wus* allele corresponds to *wus-4* (provided by Martin Hobe and Rüdiger Simon); details on the *arr3 arr4 arr5 arr6 arr7 arr8 arr9* septuple mutant are available in Supplementary Information.

Microarray experiments. Affymetrix Ath1 microarrays were hybridized as described²⁹ in duplicates using RNA from pools of 20 plants for each replicate. Expression estimates were calculated by gcRMA (ref. 30) and statistical testing for differential expression was performed with LogitT (ref. 20).

Quantitative real-time RT-PCR. qRT-PCR was performed as described²⁹ with the use of either SYBR-green or Taq-Man probes (Fig. 1d). Experiments were performed in triplicates from RNA of pooled tissue. Amplification of *TUBULIN* served as control. Oligonucleotides are listed in Supplementary Table 2.

In situ hybridization. *In situ* hybridization was performed in accordance with standard protocols, with the addition of 10% poly(vinyl alcohol) (molecular mass 70–100 kDa) to the staining solution.

ChIP. Genomic fragments were analysed by real-time qRT-PCR in triplicates. Unrelated sequences in the experimental tissue and *ARR7* sequences in leaves, where WUS should not be present, served as controls. A detailed protocol is available as Supplementary Information.

EMSA. EMSA was performed as described in ref. 17.

Transgenes. Complementary DNAs flanked by the *AlcA* promoter and the OCS terminator were inserted into a pMLBART-derived binary vector, which harbours a 35S:*AlcR* cassette¹⁸. Constitutive overexpression constructs were made in pMLBART or pART27 binary vectors using a 35S promoter and an OCS terminator.

Received 4 May; accepted 30 September 2005.

- Schoof, H. *et al.* The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100**, 635–644 (2000).
- Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M. & Simon, R. Dependence of stem cell fate in *Arabidopsis* on a feedback loop regulated by *CLV3* activity. *Science* **289**, 617–619 (2000).
- Leyser, O. Regulation of shoot branching by auxin. *Trends Plant Sci.* **8**, 541–545 (2003).
- Kiba, T. *et al.* The type-A response regulator, *ARR15*, acts as a negative regulator in the cytokinin-mediated signal transduction in *Arabidopsis thaliana*. *Plant Cell Physiol.* **44**, 868–874 (2003).
- To, J. P. *et al.* Type-A *Arabidopsis* response regulators are partially redundant negative regulators of cytokinin signalling. *Plant Cell* **16**, 658–671 (2004).
- Giulini, A., Wang, J. & Jackson, D. Control of phyllotaxy by the cytokinin-inducible response regulator homologue *ABPHYL1*. *Nature* **430**, 1031–1034 (2004).
- Laux, T., Mayer, K. F., Berger, J. & Jurgens, G. The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development* **122**, 87–96 (1996).
- Long, J. A., Moan, E. I., Medford, J. I. & Barton, M. K. A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene *Arabidopsis*. *Nature* **379**, 66–69 (1996).
- Clark, S. E., Running, M. P. & Meyerowitz, E. M. *CLAVATA1*, a regulator of meristem and flower development in *Arabidopsis*. *Development* **119**, 397–418 (1993).

- Clark, S. E., Running, M. P. & Meyerowitz, E. M. *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same processes as *CLAVATA1*. *Development* **121**, 2057–2067 (1995).
- Kayes, J. M. & Clark, S. E. *CLAVATA2*, a regulator of meristem and organ development in *Arabidopsis*. *Development* **125**, 3843–3851 (1998).
- Mayer, K. F. X. *et al.* Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95**, 805–815 (1998).
- Fletcher, J. C., Brand, U., Running, M. P., Simon, R. & Meyerowitz, E. M. Signaling of cell fate decisions by *CLAVATA3* in *Arabidopsis* shoot meristems. *Science* **283**, 1911–1914 (1999).
- Rojo, E., Sharma, V. K., Kovaleva, V., Raikhel, N. V. & Fletcher, J. C. *CLV3* is localized to the extracellular space, where it activates the *Arabidopsis* *CLAVATA* stem cell signalling pathway. *Plant Cell* **14**, 969–977 (2002).
- Clark, S. E., Williams, R. W. & Meyerowitz, E. M. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**, 575–585 (1997).
- Jeong, S., Trotochaud, A. E. & Clark, S. E. The *Arabidopsis* *CLAVATA2* gene encodes a receptor-like protein required for the stability of the *CLAVATA1* receptor-like kinase. *Plant Cell* **11**, 1925–1934 (1999).
- Lohmann, J. U. *et al.* A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell* **105**, 793–803 (2001).
- Roslan, H. A. *et al.* Characterization of the ethanol-inducible *alc* gene-expression system in *Arabidopsis thaliana*. *Plant J.* **28**, 225–235 (2001).
- Lenhard, M., Bohnert, A., Jürgens, G. & Laux, T. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* **105**, 805–814 (2001).
- Lemon, W. J., Liyanarachchi, S. & You, M. A high performance test of differential gene expression for oligonucleotide arrays. *Genome Biol.* **4**, R67 (2003).
- D'Agostino, I. B., Deruere, J. & Kieber, J. J. Characterization of the response of the *Arabidopsis* response regulator gene family to cytokinin. *Plant Physiol.* **124**, 1706–1717 (2000).
- D'Agostino, I. B. & Kieber, J. J. Phosphorelay signal transduction: the emerging family of plant response regulators. *Trends Biochem. Sci.* **24**, 452–456 (1999).
- Skoog, F. & Miller, C. O. Chemical regulation of growth and organ formation in plant tissues cultured *in vitro*. *Symp. Soc. Exp. Biol.* **54**, 118–130 (1957).
- Hwang, I. & Sheen, J. Two-component circuitry in *Arabidopsis* cytokinin signal transduction. *Nature* **413**, 383–389 (2001).
- Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nature Genet.* **37**, 501–506 (2005).
- Brand, U., Grunewald, M., Hobe, M. & Simon, R. Regulation of *CLV3* expression by two homeobox genes in *Arabidopsis*. *Plant Physiol.* **129**, 565–575 (2002).
- Lenhard, M., Jurgens, G. & Laux, T. The *WUSCHEL* and *SHOOTMERISTEMLESS* genes fulfil complementary roles in *Arabidopsis* shoot meristem regulation. *Development* **129**, 3195–3206 (2002).
- Hass, C. *et al.* The response regulator 2 mediates ethylene signalling and hormone signal integration in *Arabidopsis*. *EMBO J.* **23**, 3290–3302 (2004).
- Schmid, M. *et al.* Dissection of floral induction pathways using global expression analysis. *Development* **130**, 6001–6012 (2003).
- Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M. & Spencer, F. A. A model based background adjustment for oligonucleotide expression arrays. Working Paper 1 (Dept of Biostatistics Working Papers, Johns Hopkins Univ., Baltimore, Maryland, 2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Greenland for providing the *AlcA* system, J. Palatnik for sharing unpublished results, R. Schwab for establishing the *in situ* protocol, R. Chen for preparing the WUS antiserum, K. Harter and D. Weigel for discussion, and K. Bomblies, I. Lohmann, M. Schmid, J. Palatnik and D. Weigel for reading the manuscript critically. This work was supported by a Career Development Award of the International Human Frontier Science Program (HFSP) Organization (J.U.L.), a Ph.D. fellowship of the Cusanuswerk (W.B.), grants from the NSF and the NIH (J.J.K.) and the Max Planck Society (J.U.L.).

Author Contributions A.L. performed *in situ* hybridizations and qRT-PCRs, constructed reporter genes, the mutated alleles of *ARR7* and performed electron microscopy; J.P.C.T. and J.J.K. generated and analysed the *arr* double and septuple mutants; W.B. performed the ChIP experiments; S.S. generated constructs and transgenic lines of *ARR* genes; A.K. generated *AlcA::CLV3* plants; M.D. performed qRT-PCRs; and J.U.L. carried out the microarray experiment and analysis, performed gel-shifts and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Microarray data have been deposited at ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MEXP-432. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.U.L. (jlohmam@tuebingen.mpg.de).

LETTERS

Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin

Toru Hirota^{1†}, Jesse J. Lipp¹, Ban-Hock Toh² & Jan-Michael Peters¹

Histones are subject to numerous post-translational modifications¹. Some of these 'epigenetic' marks recruit proteins that modulate chromatin structure. For example, heterochromatin protein 1 (HP1) binds to histone H3 when its lysine 9 residue has been tri-methylated by the methyltransferase Suv39h (refs 2–6). During mitosis, H3 is also phosphorylated by the kinase Aurora B⁷. Although H3 phosphorylation is a hallmark of mitosis, its function remains mysterious. It has been proposed that histone phosphorylation controls the binding of proteins to chromatin⁸, but any such mechanisms are unknown. Here we show that antibodies against mitotic chromosomal antigens that are associated with human autoimmune diseases⁹ specifically recognize H3 molecules that are modified by both tri-methylation of lysine 9 and phosphorylation of serine 10 (H3K9me3S10ph). The generation of H3K9me3S10ph depends on Suv39h and Aurora B, and occurs at pericentric heterochromatin during mitosis in different eukaryotes. Most HP1 typically dissociates from chromosomes during mitosis^{10–12}, but if phosphorylation of H3 serine 10 is inhibited, HP1 remains chromosome-bound throughout mitosis. H3 phosphorylation by Aurora B is therefore part of a 'methyl/phos switch' mechanism⁸ that displaces HP1 and perhaps other proteins from mitotic heterochromatin.

Most eukaryotic chromosomes contain specialized gene-poor domains, called centromeres, which have at least two essential functions in chromosome segregation: they nucleate the assembly of kinetochores, the structures that are captured by spindle microtubules, and they bind and protect cohesin complexes that mediate sister chromatid cohesion, thereby enabling chromosome bi-orientation in mitosis^{13–15}. In many organisms, centromeres are characterized by the presence of pericentric heterochromatin—short, highly repetitive sequence elements organized in domains that stain intensely with DAPI and appear dense in electron micrographs of interphase nuclei^{13,15}. In these domains, methyltransferases modify specific lysine residues on histone H3, creating binding sites for proteins such as HP1 (Swi6 in fission yeast) and Polycomb^{2–6,16}. These proteins repress gene expression by modulating chromatin structure¹, and in fission yeast HP1/Swi6 is also required for enrichment of cohesin at centromeres^{17,18}.

Apart from HP1 and cohesin, only a few molecules are known to have centromeric functions. These include the Aurora B chromosome passenger complex, the centromere-specific histone CENP-A, the cohesin protector Sgo1, and components of the RNA interference machinery^{13–15,19}. To identify molecules or modifications that are located at centromeres, we searched for human autoimmune sera that specifically react with mitotic centromeres in immunofluorescence microscopy experiments. This approach has previously led to the identification of several centromere proteins, including CENP-A²⁰. Out of 32,000 human sera screened for anti-nuclear

antibodies, we identified six that reacted specifically with mitotic chromosomes (Supplementary Table and Supplementary Fig. 1; ref. 9). These antibodies were associated with different diseases, including discoid lupus erythematosus, Sjögren's syndrome, polymyalgia rheumatica and rheumatoid arthritis, indicating that the occurrence of antibodies against mitotic chromosomal antigens (MCAs) is not restricted to a specific syndrome. We characterized three of the six sera (MCA1, -2 and -6), because these specifically stained centromeres from prophase until anaphase (Fig. 1a) and because their epitopes could be identified unequivocally (see below and Supplementary Table). Data shown here are for MCA1, but similar results were obtained with MCA2 and MCA6 (Supplementary Fig. 1).

Immunofluorescence microscopy co-localization experiments in mouse embryonic fibroblasts (MEFs) (Fig. 1a) and human HeLa cells (Fig. 1c) showed that MCA1 reacts with chromosomes during the same stages of mitosis in which H3 is phosphorylated on Ser 10 (H3S10ph), but that MCA1 stains predominantly centromeres, whereas H3S10ph antibodies also stain chromosome arms. MCA1 staining was seen in a discrete domain adjacent to the kinetochores that stains intensely with DAPI and antibodies against H3K9me3, indicating that MCA1 reacts with a component of pericentric heterochromatin (Fig. 1c). In immunoblot analyses of HeLa cell lysates, MCA1 recognized a mitosis-specific 17 kDa protein band that co-migrated with H3 (Fig. 1b). As Aurora B is known to modify H3 in mitosis⁷, and as MCA1 reactivity is sensitive to phosphatase treatments⁹, we asked whether generation of the MCA1 epitope is prevented by hesperadin, a compound that inhibits Aurora B²¹. No MCA1 reactivity was detected in mitotic cells treated with hesperadin (Fig. 1d), implying that MCA1 reacts with a mitotic modification on H3, the generation of which depends on Aurora B.

To identify this modification, we used synthetic H3 peptides as potential competitors of MCA1 reactivity in immunofluorescence microscopy experiments. Out of six phospho-peptides derived from the H3 amino terminus (Fig. 2a), none reduced MCA1 reactivity (Supplementary Fig. 2). As the MCA1 epitope is hesperadin-sensitive and is enriched in heterochromatin, where H3 is highly methylated, we next used peptides in which serine/threonine residues are phosphorylated, and in which adjacent lysine residues are mono-, di-, or tri-methylated. Of the resulting twelve peptides, only one—containing tri-methylated Lys 9 and phosphorylated Ser 10 (H3K9me3S10ph)—abolished MCA1 staining (Fig. 2b). In contrast, peptides that were tri-methylated on Lys 9 but were not phosphorylated on Ser 10 did not reduce MCA1 reactivity (Fig. 2b). Dot-blot experiments confirmed that MCA1 reacted specifically with H3K9me3S10ph peptides, but not with otherwise identical peptides that were either not tri-methylated on Lys 9 or not phosphorylated on Ser 10 (Fig. 2c).

¹Research Institute of Molecular Pathology, Dr. Bohrgasse 7, A-1030 Vienna, Austria. ²Department of Immunology, Monash University Medical School, Commercial Road, Prahran, Victoria 3181, Australia. [†]Present address: Department of Experimental Pathology, Cancer Institute, Japanese Foundation for Cancer Research, 3-10-6 Ariake, Koto-ku, Tokyo 135-8550, Japan.

The finding that MCA1 reacts with H3K9me3S10ph is consistent with our observations that the MCA1 epitope is only generated in mitosis, depends on Aurora B activity and is enriched at pericentric heterochromatin. It further predicts that formation of the MCA1 epitope depends on Suv39h. Indeed, no MCA1 reactivity was observed in mitotic fibroblasts derived from Suv39h-deficient mice, although MCA1 did stain chromosomes in wild-type (Suv39h-proficient) mouse cells (Fig. 1d). Our data therefore indicate that the pericentric heterochromatin of mitotic chromosomes contains H3 molecules that are doubly modified, by tri-methylation of Lys 9 and phosphorylation of Ser 10. Earlier *in vitro* studies had implied that these two modifications might only occur in a mutually exclusive manner²², but evidence for the existence of H3K9me3S10ph has recently also been obtained by mass spectrometry⁸. As we observed that MCA1 also reacts with mitotic chromosomes in other species, including *Drosophila* (Supplementary Fig. 3), it is possible that the occurrence of H3K9me3S10ph is widespread and has been conserved during evolution.

H3S10 phosphorylation by Aurora B is a hallmark of mitosis, and

as such is widely used as a mitotic marker. However, the molecular function of this modification remains unknown. H3S10 phosphorylation has been implicated in chromosome condensation in *Drosophila*²³, but in human cells chromosomes can condense apparently normally without this modification²¹. Mutation of H3 Ser 10 to alanine does not lead to detectable phenotypes in budding yeast, but the same mutation causes chromosome segregation defects in fission yeast²⁴ and *Tetrahymena*²⁵, the molecular cause of which is unknown. It has recently been proposed that Ser 10 phosphorylation can prevent binding of HP1 to the adjacent tri-methylated Lys 9 residue of H3 (ref. 8). To test this hypothesis, and to gain insight into the molecular functions of H3S10 phosphorylation, we analysed the role of Aurora B in controlling the association of HP1 with heterochromatin.

In interphase cells, antibodies against the α -isoform of HP1 stain nuclear speckles that represent heterochromatin domains. We confirmed the observation that these signals gradually disappear during prophase and are hardly detectable by prometaphase^{10–12}, indicating that HP1 α dissociates from chromosomes during mitosis

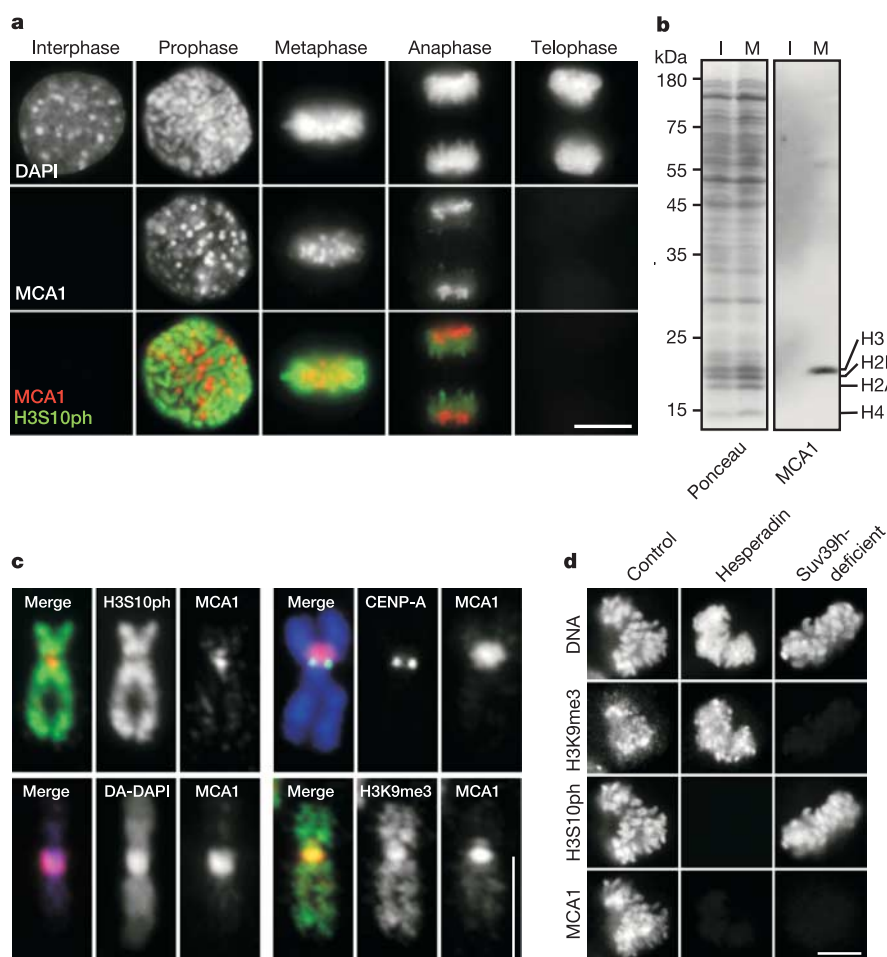


Figure 1 | Characterization of the MCA1 serum. **a**, Immunofluorescence microscopy analysis of MCA1. Logarithmically proliferating MEFs were fixed and stained with MCA1 serum (red) and anti-H3S10ph (green) antibodies. DNA was counterstained with DAPI. Representative cells at different stages of the cell cycle are shown. Scale bar, 10 μ m. **b**, MCA1 recognizes a mitosis-specific modification on H3. Total HeLa cell extracts from cells arrested in S phase by hydroxyurea treatment (I) or arrested in mitosis by nocodazole treatment (M) were resolved by 14% SDS-PAGE, transferred to PVDF membrane and stained with Ponceau S (left) or probed with MCA1 serum (right). Bands corresponding to core histones (H2A, H2B, H3 and H4) are indicated. **c**, Enrichment of MCA1 epitope at pericentric heterochromatin of metaphase chromosomes. Chromosomes

from HeLa cells were co-stained with MCA1 serum and antibodies to either H3S10ph (upper left) or H3K9me3 (lower right). Heterochromatic DNA was visualized by distamycin (DA) and DAPI staining²⁹ (lower left). Kinetochore were visualized by stable expression of CENP-A fused (at its N terminus) to enhanced green fluorescent protein (EGFP) (upper right). Scale bar, 5 μ m. **d**, Generation of the MCA1 epitope requires the activity of Aurora B and Suv39h. Wild-type MEFs treated with (middle) or without (left) 250 nM hesperadin, or MEFs derived from Suv39h-deficient mice (Suv39h dn, right panels), were triple-stained using antibodies against H3S10ph, H3K9me3 and MCA1. DNA was stained with TOTO3-iodide. Representative prometaphase cells are shown. Scale bar, 10 μ m.

(Supplementary Fig. 4a). Notably, this decline in the HP1 α signal coincided with the appearance of MCA1 staining in early mitosis (Supplementary Fig. 4a). However, when cells were treated with hesperadin, HP1 α remained enriched at pericentric heterochromatin throughout mitosis (Fig. 3a, b and Supplementary Fig. 4b). We obtained consistent results in biochemical experiments in which subcellular fractions were analysed by immunoblotting. In mitotic control cells, a portion of HP1 α could be detected in cytoplasmic fractions, but almost all HP1 α was found in chromosome fractions when hesperadin-treated cells were processed in the same way (Fig. 3c). Similar observations were made in immunofluorescence microscopy experiments for HP1 β and HP1 γ , for which dissociation from mitotic chromosomes was also reduced by hesperadin treatment (Supplementary Fig. 5).

These observations indicate that a hesperadin-sensitive kinase is required for dissociation of the different isoforms of HP1 from heterochromatin during mitosis. To test whether this kinase is Aurora B, we performed RNA interference experiments (Supplementary Fig. 6) and analysed the distribution of HP1 α by immunofluorescence microscopy (Fig. 3d). As predicted, depletion of Aurora B abolished MCA1 staining and prevented dissociation of HP1 α from mitotic chromosomes, whereas reduction of Aurora A did not interfere with HP1 α behaviour (Fig. 3d).

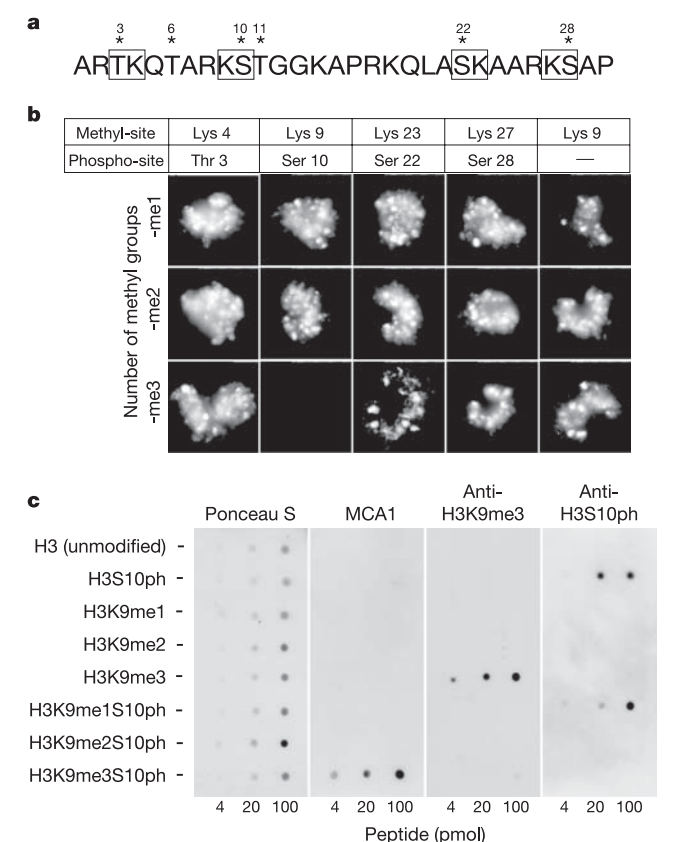


Figure 2 | Identification of the MCA1 epitope. **a**, Amino acid sequence of the histone H3 N terminus, which is conserved from yeast to humans. Candidate sites for phosphorylation are marked with asterisks. Possible double modification sites for phosphorylation and methylation are highlighted with boxes. **b**, Peptide competition of MCA1 staining. The indicated H3 peptides (30 μ M) were preincubated with antibodies for 60 min and then used in immunofluorescence microscopy analysis. **c**, Dot-blot analysis of H3 peptides. Three dilutions (100, 20 and 4 pmol) of peptides with the indicated modifications were spotted onto a membrane and stained with Ponceau S (left). The membrane was then probed with MCA1 serum or with antibodies against H3K9me3 or H3S10ph as indicated.

These results show that Aurora B is required for the dissociation of HP1 proteins from pericentric heterochromatin during mitosis. To address whether Aurora B mediates this function by phosphorylating H3 on Ser 10, we analysed the ability of recombinant HP1 α to bind to

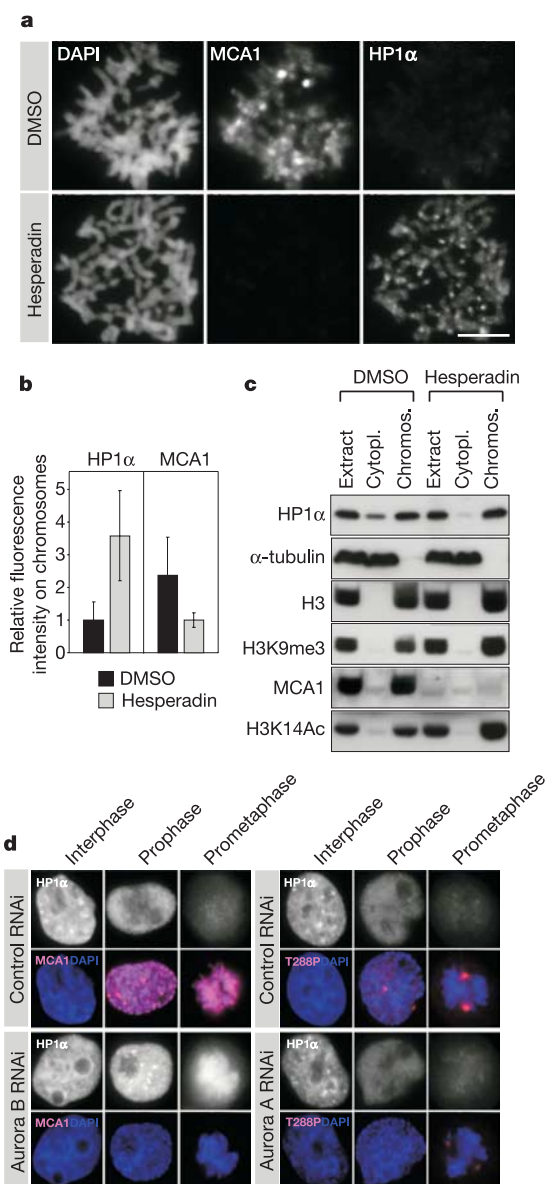


Figure 3 | HP1 α remains associated with mitotic chromosomes if Aurora B is inhibited. **a**, Localization of HP1 α and MCA1 in chromosome spreads. Nocodazole-arrested mitotic cells were treated for 30 min with 0.1% DMSO with or without 100 nM hesperadin. Scale bar, 10 μ m. **b**, Quantifying the fluorescence intensities of MCA1 and HP1 α on chromosomes. Images obtained as in **a** were quantified, and the resulting relative values (mean $\pm$ s.e.m.) are shown in histograms ($n = 30$ for both DMSO and hesperadin). **c**, Loss of HP1 α from cytoplasmic fractions after Aurora B inhibition. Mitotic HeLa cells were incubated for 30 min in the presence or absence of 100 nM hesperadin. Total cell extracts were fractionated into cytoplasmic fractions (cytopl) and chromosome-enriched fractions (chromos) and analysed by immunoblotting with the indicated antibodies. **d**, Depletion of Aurora B, but not of Aurora A, causes persistence of HP1 α on mitotic chromosomes. Synchronous HeLa cells were transfected with siRNAs against Aurora A, Aurora B or to a control protein, and were stained with anti-HP1 α antibodies (top panels). Depletion of Aurora B or Aurora A was confirmed by western blotting (Supplementary Fig. 6) and by the reduced reactivity of MCA1 and Aurora A phospho-Thr 288 antibodies, respectively (red, bottom panels).

H3K9me3 peptides in the presence or absence of Ser 10 phosphorylation. HP1 α bound much less efficiently to H3K9me3S10ph than to H3K9me3 peptides (Fig. 4a). This difference was abolished if H3K9me3S10ph peptides were pre-incubated with phosphatase (Fig. 4b), indicating that phosphorylation of Ser 10 is the cause of reduced HP1 α binding to H3K9me3S10ph. To test whether Aurora B activity is sufficient to prevent binding of HP1 α to H3K9me3, we pre-incubated these peptides with ATP and recombinant active Aurora B. This treatment reduced HP1 α binding (Fig. 4c). The inhibitory effect required Aurora B kinase activity, as hesperadin abolished this effect. Incubation of H3K9me3 peptides with Aurora B also enabled recognition of these peptides by MCA1 antibodies, indicating that Aurora B is able to phosphorylate Ser 10. To address whether phosphorylation of Ser 10 is required for the inhibitory effect of Aurora B on HP1 α binding, we analysed the binding of HP1 α to H3K9me3 peptides in which Ser 10 was replaced by alanine (H3K9me3S10A). HP1 α bound to these peptides, although less well than to peptides containing Ser 10, and this binding was not reduced by treating the peptides with Aurora B.

Our results indicate that phosphorylation of Ser 10 by Aurora-B is necessary and sufficient for the dissociation of HP1 α from H3, at

least *in vitro*. However, it has recently been reported that acetylation of Lys 14 is required to prevent binding of HP1 α to H3 (ref. 26). We were unable to detect a reduction in HP1 α binding to H3 peptides in which Lys 14 was acetylated, but phosphorylation of Ser 10 prevented HP1 α binding independent of Lys 14 acetylation (Fig. 4a). We can not exclude the possibility that Lys 14 acetylation has a role in controlling HP1 α behaviour *in vivo*, but our observations that hesperadin treatment and Aurora B depletion are sufficient to prevent HP1 α dissociation from mitotic chromosomes (Fig. 3 and Supplementary Fig. 4) are more easily explained by the proposal that H3 phosphorylation on Ser 10 determines whether HP1 α remains bound to chromatin.

We therefore conclude that Aurora B mediates the dissociation of HP1 proteins from heterochromatin by phosphorylating H3 on Ser 10 (Fig. 4d). The negative charge of phosphorylated Ser 10 might repel the negatively charged Glu 52 residue in the chromodomain groove of HP1 α . It will be of interest to determine whether phosphorylation of H3S28 by Aurora B²⁷ is responsible for the mitotic dissociation of Polycomb from the adjacent methylated Lys 27 residue¹⁶.

Our data provide molecular insight into the function of H3 Ser 10 phosphorylation. Notably, inhibition or depletion of Aurora B causes a number of defects in chromosome structure and behaviour, including reduction of cohesin dissociation from chromosome arms, lack of a primary constriction at centromeres, abnormally fuzzy appearance of chromosomes in hypotonic buffers, and syntelic attachments of sister kinetochores to microtubules from the same spindle pole (refs 21, 28 and unpublished observations). It will be interesting to determine whether any of these defects are caused by the abnormal persistence of HP1 on mitotic chromosomes. For example, as HP1/Swi6 positively regulates the association of cohesin with chromosomes in fission yeast^{17,18}, it is possible that HP1 dissociation in mitosis is required for the removal of cohesin from chromosome arms in early mitosis. Likewise, it will be important to understand whether the behaviour of HP1 is causally related to chromosome segregation defects that have been observed in fission yeast²⁴ and *Tetrahymena*²⁵ mutants in which H3 cannot be phosphorylated on Ser 10, and whether H3S10 phosphorylation may have additional, as yet unknown functions.

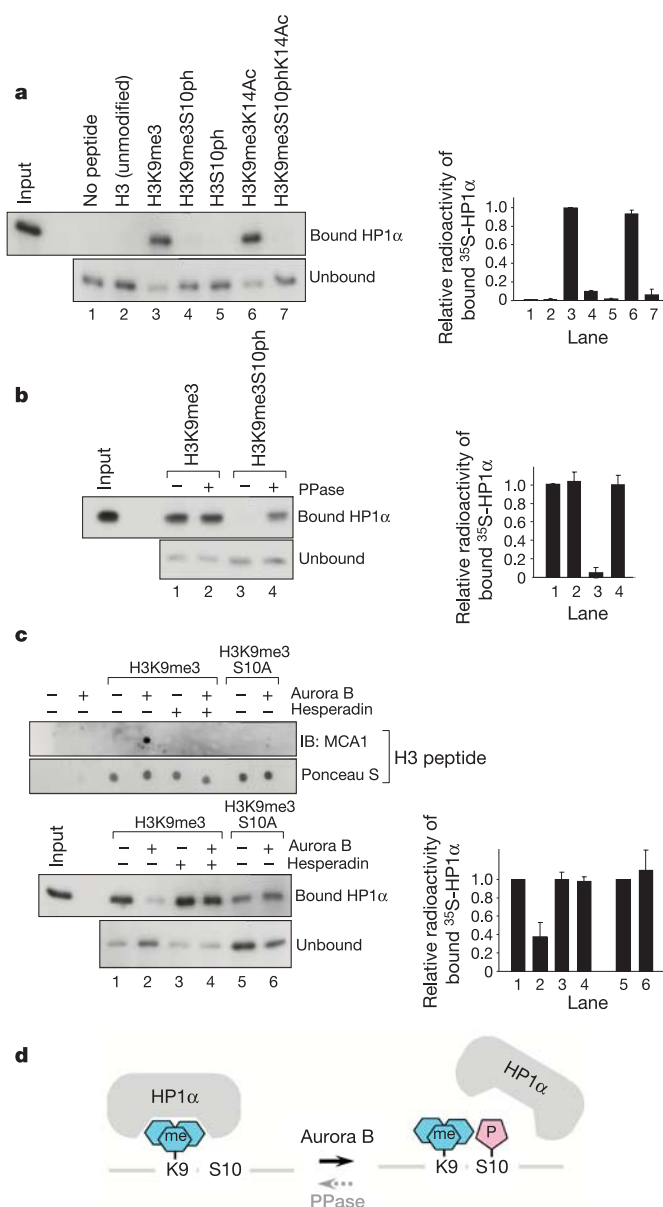


Figure 4 | H3 Ser 10 phosphorylation is necessary and sufficient to prevent binding of HP1 α to H3 peptides. **a**, HP1 α does not bind stably to H3K9me3S10ph peptides. *In vitro* translated Myc-tagged ³⁵S-HP1 α was incubated with a series of modified peptide-coupled beads, and HP1 α in the bound and unbound fractions was analysed by anti-Myc immunoblotting (left). The amounts of ³⁵S-HP1 α bound to beads were measured, and mean values ($\pm$ s.e.m.) from three independent experiments are shown (right). The amount of HP1 α bound to H3K9me3 peptides (lane 3) was set to 1.0. **b**, Phosphatase treatment of H3K9me3S10ph peptides restores binding to HP1 α . H3K9me3 or H3K9me3S10ph peptide-coupled beads were incubated in the presence or absence of alkaline phosphatase (PPase), and subsequently used in HP1 α binding assays (left). HP1 α binding was quantified as in **a** (mean $\pm$ s.e.m.; $n = 3$), and the amount of HP1 α in lane 1 was set to 1.0 (right). **c**, Aurora-B-mediated phosphorylation of Ser 10 on H3K9me3 peptides prevents binding to HP1 α . H3K9me3 or H3K9me3S10A peptides were incubated in the presence or absence of recombinant Aurora B with or without hesperadin, and phosphorylation of Ser 10 was detected by MCA1 antibodies (upper left). Bead-coupled H3 peptides were treated in the same way and used in the HP1 α binding assay (lower left). Because HP1 α bound less effectively to H3K9me3S10A than to H3K9me3, twice the amount of HP1 α was used (lanes 5 and 6). HP1 α binding was quantified as in **a** (mean $\pm$ s.e.m.; $n = 3$). The amount of HP1 α in lanes 1 and 5 is set to 1.0 as a reference for lanes 2–4 and lane 6, respectively. **d**, A model of the 'methyl/phos switch' mechanism that controls the association of HP1 with H3. Mitotic H3 Ser 10 phosphorylation mediated by Aurora B is sufficient to prevent binding of the HP1 α chromodomain to H3K9me3. At the end of mitosis, protein phosphatases can revert the effect of Aurora B by dephosphorylating Ser 10.

METHODS

Immunofluorescence microscopy. Cells grown on coverslips were fixed with 4% paraformaldehyde in phosphate buffer pH 7.4 for 20 min, permeabilized with 0.2% Triton X-100 in phosphate-buffered saline (PBS) for 5 min, and incubated with 3% bovine serum albumin (BSA). The following primary antibodies were used: anti-HP1 α monoclonal (Euromedex), anti-H3S10ph (clone 6G3) and anti-Aurora-A T288P (Cell Signalling Technologies). Anti-H3K9me3 antibodies were a gift from T. Jenuwein. All washes were performed using 0.01% Triton X-100 in PBS, to which 0.3% BSA was added for antibody dilutions. Images were captured on a Zeiss Axioplan 2 microscope as described²⁹.

For chromosome spreads, mitotic cells were collected by shake-off, resuspended in 2 ml medium, mixed with 3 ml tap water, incubated for 5 min and then centrifuged onto ethanol-prewashed slideglasses at 250g for 5 min in a cytocentrifuge (Cytospin 4, Shandon). Cells were then pre-extracted with 0.1% Triton X-100 in PBS for 2 min, washed in PBS for 3 min, fixed with 2% paraformaldehyde in phosphate buffer for 15 min, and processed for immunofluorescence microscopy analysis. Fluorescence intensities were quantified using Image-J software.

Dot-blot analysis. Two-microlitre serial dilutions of peptides in PBS were spotted onto 0.45- μ m pore-sized PVDF membrane (Immobilon-P, Millipore) and fixed with 2% formaldehyde/PBS for 10 min.

HP1 binding assays. Synthetic peptides (500 nmol) were covalently coupled through carboxy-terminal cysteine residues to 5 mg of activated POROS matrix (Applied Biosystems) in HEPES buffer (pH 7.3) plus 0.5 mM EDTA. *In vitro* pull-down assays were carried out as described⁴. Briefly, 10 μ l of peptide-coupled beads were incubated for 1 h with 1 μ l of ³⁵S-labelled *in vitro* translated HP1 α protein (TNT rabbit reticulocyte lysate system, Promega) in a volume of 100 μ l of binding buffer. The input, bound and unbound proteins were separated by 14% SDS-PAGE and visualized by anti-myc epitope immunoblotting and autoradiography.

For phosphatase treatment, beads were equilibrated in buffer containing 50 mM Tris (pH 8.5) and 5 mM MgCl₂, and incubated with 20 U alkaline phosphatase (Roche) for 30 min at 37 °C. For kinase reactions, beads were equilibrated with a kinase buffer consisting of 20 mM Tris (pH 7.4), 5 mM MgCl₂, 1 mM EGTA and 150 mM NaCl, and incubated for 30 min at 30 °C in a total volume of 30 μ l supplemented with 50 μ M ATP, 1 μ M purified Aurora B (residues 60–361) bound to its subunit inner centromere protein (INCENP, residues 790–856) (ref. 30) and either 1 μ M hesperadin or an equivalent concentration of dimethylsulphoxide (DMSO). Beads were then washed extensively with the washing buffer⁴ before being used in pull-down assays.

Cell extract fractionation. Nocodazole-arrested mitotic HeLa cells were collected by shake-off and incubated for 30 min at 37 °C in the presence or absence of 100 nM hesperadin. After washing with PBS, cells were lysed on ice for 10 min by Dounce homogenization in a buffer consisting of 10 mM HEPES (pH 7.9), 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10% glycerol, 1 mM dithiothreitol, 0.25% TritonX-100 and a protease inhibitor cocktail (Roche). Chromosome-enriched fractions were collected by low-speed centrifugation at 1300g for 5 min, and were washed twice with the same buffer. Cytoplasmic fractions were prepared from the supernatant by further centrifugation at 16000g for 20 min.

Received 2 August; accepted 23 September 2005.

Published online 12 October 2005.

- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
- Jacobs, S. A. *et al.* Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J.* **20**, 5232–5241 (2001).
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
- Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
- Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
- Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291 (2000).
- Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
- Gitlits, V. M., Macaulay, S. L., Toh, B. H. & Sentry, J. W. Novel human

autoantibodies to phosphopeptides on mitotic chromosomal autoantigens (MCAs). *J. Invest. Med.* **48**, 172–182 (2000).

- Murzina, N., Verreault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
- Sugimoto, K., Tasaka, H. & Dotsu, M. Molecular behaviour in living mitotic cells of human centromere heterochromatin protein HPL α ectopically expressed as a fusion to red fluorescent protein. *Cell Struct. Funct.* **26**, 705–718 (2001).
- Schmiedeberg, L., Weisshart, K., Diekmann, S., Meyer Zu Hoerste, G. & Hemmerich, P. High- and low-mobility populations of HP1 in heterochromatin of mammalian cells. *Mol. Biol. Cell* **15**, 2819–2833 (2004).
- Sullivan, B. A., Blower, M. D. & Karpen, G. H. Determining centromere identity: cyclical stories and forking paths. *Nature Rev. Genet.* **2**, 584–596 (2001).
- Mellone, B. G. & Allshire, R. C. Stretching it: putting the CEN(P-A) in centromere. *Curr. Opin. Genet. Dev.* **13**, 191–198 (2003).
- Amor, D. J., Kalitsis, P., Sumer, H. & Choo, K. H. Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol.* **14**, 359–368 (2004).
- Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
- Bernard, P. *et al.* Requirement of heterochromatin for cohesion at centromeres. *Science* **294**, 2539–2542 (2001).
- Nonaka, N. *et al.* Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature Cell Biol.* **4**, 89–93 (2002).
- Grewal, S. I. & Rice, J. C. Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell Biol.* **16**, 230–238 (2004).
- Earnshaw, W., Bordwell, B., Marino, C. & Rothfield, N. Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies. *J. Clin. Invest.* **77**, 426–430 (1986).
- Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161**, 281–294 (2003).
- Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
- Giet, R. & Glover, D. M. *Drosophila* aurora B kinase is required for histone H3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis. *J. Cell Biol.* **152**, 669–682 (2001).
- Mellone, B. G. *et al.* Centromere silencing and function in fission yeast is governed by the amino terminus of histone H3. *Curr. Biol.* **13**, 1748–1757 (2003).
- Wei, Y., Yu, L., Bowen, J., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**, 99–109 (1999).
- Mateescu, B., England, P., Halgand, F., Yaniv, M. & Muchardt, C. Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. *EMBO Rep.* **5**, 490–496 (2004).
- Goto, H., Yasui, Y., Nigg, E. A. & Inagaki, M. Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes Cells* **7**, 11–17 (2002).
- Gimenez-Abian, J. F. *et al.* Regulation of sister chromatid cohesion between chromosome arms. *Curr. Biol.* **14**, 1187–1193 (2004).
- Hirota, T., Gerlich, D., Koch, B., Ellenberg, J. & Peters, J. M. Distinct functions of condensin I and II in mitotic chromosome assembly. *J. Cell Sci.* **117**, 6435–6445 (2004).
- Sessa, F. *et al.* Mechanism of Aurora B activation by INCENP and inhibition by hesperadin. *Mol. Cell* **18**, 379–391 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to K. Mechtler and M. Madalinski for peptide synthesis, to S. Opravil and T. Jenuwein for H3K9me3 antibodies and Suv39h-deficient MEFs, to A. Musacchio for purified Aurora B-INCENP complex, to H. Saya for Aurora A antibodies, and to W. Pollock for collecting MCA sera sent to Gribbles Pathology. T.H. acknowledges a fellowship from the Japanese Society for the Promotion of the Science (JSPS). Research in the laboratory of J.-M.P. is supported by Boehringer Ingelheim, Wiener Wirtschaftsfoerderungsfonds (WWFF), the Sixth Framework Programme of the European Union via the Integrated Project Mitochek, the Austrian Science Fund and the European Molecular Biology Organization (EMBO).

Author Contributions B.-H.T. identified MCA sera. T.H. and J.-M.P. conceived and designed the experiments. T.H. and J.J.L. performed the experiments and analysed the data. T.H. and J.-M.P. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.-M.P. (peters@imp.univie.ac.at).

Double chromodomains cooperate to recognize the methylated histone H3 tail

John F. Flanagan¹, Li-Zhi Mi^{2†}, Maksymilian Chruszcz³, Marcin Cymborowski³, Katrina L. Clines¹, Youngchang Kim⁴, Wladek Minor³, Fraydoon Rastinejad^{1,2} & Sepideh Khorasanizadeh¹

Chromodomains are modules implicated in the recognition of lysine-methylated histone tails and nucleic acids^{1,2}. CHD (for chromo-ATPase/helicase-DNA-binding) proteins regulate ATP-dependent nucleosome assembly and mobilization through their conserved double chromodomains and SWI2/SNF2 helicase/ATPase domain^{3–5}. The *Drosophila* CHD1 localizes to the interbands and puffs of the polytene chromosomes, which are classic sites of transcriptional activity⁶. Other CHD isoforms (CHD3/4 or Mi-2) are important for nucleosome remodelling in histone deacetylase complexes^{7,8}. Deletion of chromodomains impairs nucleosome binding and remodelling by CHD proteins⁴. Here we describe the structure of the tandem arrangement of the human CHD1 chromodomains, and its interactions with histone tails. Unlike HP1 and Polycomb proteins that use single chromodomains to bind to their respective methylated histone H3 tails, the two chromodomains of CHD1 cooperate to interact with one methylated H3 tail. We show that the human CHD1 double chromodomains target the lysine 4-methylated histone H3 tail (H3K4me), a hallmark of active chromatin⁹. Methylammonium recognition involves two aromatic residues, not the three-residue aromatic cage used by chromodomains of HP1 and Polycomb proteins^{10–13}. Furthermore, unique inserts within chromodomain 1 of CHD1 block the expected site of H3 tail binding seen in HP1 and Polycomb, instead directing H3 binding to a groove at the inter-chromodomain junction.

We previously showed that the highly homologous chromodomains of HP1 and Polycomb proteins are discriminatory for binding to the methylated lysine 9 and lysine 27 in histone H3 (H3K9me and H3K27me), which are modifications associated with constitutive and facultative heterochromatin, respectively¹². Unlike HP1 and Polycomb, the CHD1 protein contains tandem chromodomains and is present in sites of transcriptional activity in *Drosophila*⁶. In yeast, CHD1 was identified in the SAGA/SLIK histone acetyltransferase complexes, and chromodomain 2 has been suggested to be responsible for H3K4me binding¹⁴. We therefore postulated that the human CHD1 chromodomains interact with the histone modifications associated with active chromatin; these include H3K4me, H3K36me and H3K79me as well as multiple lysine acetylations in the histone H3 tail.

We prepared a recombinant construct corresponding to the double chromodomain region (Fig. 1a) and used it for structural and biochemical studies with post-translationally modified histone H3 tails. The crystal structure in Fig. 1b, c shows the overall organization of the two chromodomains in the human CHD1 protein (Supplementary Table S1). Both chromodomains share their general secondary structure elements with those previously seen in HP1 and Polycomb chromodomains (Fig. 1d). The linker

segment forms a novel helix–turn–helix structure that juxtaposes the two chromodomains to form a continuous surface. A total of 350 Å² is buried at the interface of these tandem chromodomains.

To discover the specificity of the human CHD1 binding to methylated histone tails, we performed a series of fluorescence polarization-based assays with the use of synthetic peptides (Fig. 2a, and Supplementary Table S2). CHD1 binds only to the lysine 4-methylated H3 tail; for trimethylated lysine 4 (H3K4me3) the dissociation constant K_d is 5 μM and for monomethylated lysine 4 (H3K4me1) K_d is 17 μM. Importantly, CHD1 does not interact with the unmodified H3 tail or other histone codes associated with active chromatin, for example H3K36me or H3K79me. To understand the stereochemical basis for this specificity, we solved the crystal structure of the human CHD1 double chromodomains, in complex with the H3 tail containing both trimethyllysine and monomethyllysine at residue 4, at 2.40 and 2.65 Å resolutions, respectively (Supplementary Table S1). Difference maps showed electron density for bound H3 peptide (Fig. 2b). The trimethyllysine and monomethyllysine peptide complexes are essentially superimposable (Figs 2b and 3a).

The peptide was not located on either of the two chromodomain sites corresponding to the binding site on the HP1 and Polycomb chromodomains. Instead, the H3 tail was located at an acidic surface bridging chromodomains 1 and 2 (Fig. 2b, c). The unexpected cooperation between the CHD1 chromodomains to form a single site for peptide binding was further confirmed by isothermal titration calorimetry, which showed that their interaction involves the binding of one H3 tail to one double chromodomain polypeptide (Supplementary Fig. S1).

Peptide binding buries 475 Å² of surface on both chromodomains (Fig. 2b, c). This interaction involves residues 37, 64, 66 and 67 from chromodomain 1 and residues 150 and 166 from chromodomain 2, cooperating to bind residues 1–5 in the H3 tail (Fig. 2d), and residues 6 and 7 not forming direct contacts with either chromodomains. Figure 3a shows how the methylammonium of lysine 4 is recognized. The human CHD1 uses only two aromatic residues (tryptophans 64 and 67) for methyllysine recognition, not the three-residue aromatic cage seen previously in HP1 and Polycomb chromodomains (Fig. 3a, b)^{10,12}. The lack of a third aromatic residue in the formation of a π-electron cage surrounding the methylammonium seems to be compensated for by an adjacent cation–π interaction involving the peptide arginine 2 and CHD1 tryptophan 67. Figure 3c shows that mutation of either tryptophan 64 or 67 to leucine substantially reduces the binding affinity for histone H3 tail methylated on lysine 4, indicating that both tryptophan residues are required for the methylammonium recognition.

However, different members of the CHD protein family exhibit

¹Department of Biochemistry and Molecular Genetics, ²Department of Pharmacology and ³Department of Molecular Physiology and Biological Physics, University of Virginia Health System, Charlottesville, Virginia 22908, USA. ⁴Argonne National Laboratory, Biosciences Division/Structural Biology Center, Argonne, Illinois 60439, USA. [†]Present address: Department of Pathology, Harvard Medical School and The CBR Institute for Biomedical Research, Inc., 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

sequence variations at residue 67. In humans, three of the eight CHD proteins—CHD3 (Mi-2 α), CHD4 (Mi-2 β) and CHD5—have a leucine or a methionine residue instead of tryptophan 67, indicating that these isoforms might not interact with the lysine 4-methylated H3 tail. Within the subfamily CHD1, the budding yeast does not have an aromatic residue at position 67 (Supplementary Fig. S2). This indicates that unlike the CHD1 from fission yeast or metazoans, the budding yeast CHD1 would not interact with the methylated H3 tail; this was confirmed in the binding studies shown in Fig. 3c and a previous report¹⁵. Thus, these data and also our studies with the isolated chromodomain 2 region of the budding yeast CHD1 (Supplementary Table S2) disagree with previous *in vitro* studies¹⁴ indicating that chromodomain 2 of budding yeast CHD1 binds to the H3K4me peptide.

The intimate recognition of an arginine residue at position $n - 2$ relative to the methyllysine carries significance for the specificity of the human CHD1 chromodomains (Fig. 3a). No other lysine methylation site in histones contains an arginine at position $n - 2$. Because arginine 2 can be methylated by coactivator-associated arginine methyltransferase 1 (CARM1)¹⁶, we tested whether this methylation further affected the binding of the H3 tail. Methylation

of arginine 2 (Arg 2 asymmetric dimethylation; H3R2me2a) alone does not allow binding to CHD1 chromodomains; however, this modification together with methyllysine 4 (H3K4me3R2me2a) decreased the binding affinity fourfold relative to methyllysine 4 binding alone (Fig. 3d, and Supplementary Table S2). We then solved the crystal structure of the doubly modified peptide in complex with CHD1 chromodomains (Fig. 3a) and found only one significant difference from the methyllysine 4 complex that contributes to weaker binding. Methylation of arginine 2 prevented the hydrogen bonding between its side chain and the glycine 66 backbone carbonyl (Fig. 2d).

Another modification that occurs proximal to methyllysine 4 is the phosphorylation of threonine 3. Mitotic cells preserve their lysine 4 methylation patterns to maintain active chromatin states through cell division¹⁷. Threonine 3 becomes phosphorylated during mitosis by the kinase haspin, and a lysine 4-methylated H3 tail can be phosphorylated at threonine 3 by haspin kinase *in vitro*¹⁸. These findings imply that H3K4me and H3T3ph modifications coexist during mitosis. Previous studies have indicated that CHD1 is released into the cytoplasm when cells enter mitosis¹⁹, so phosphorylation no longer supports the binding of CHD1 protein *in vivo*. Threonine 3 phosphorylation is reversed during anaphase¹⁸, and this

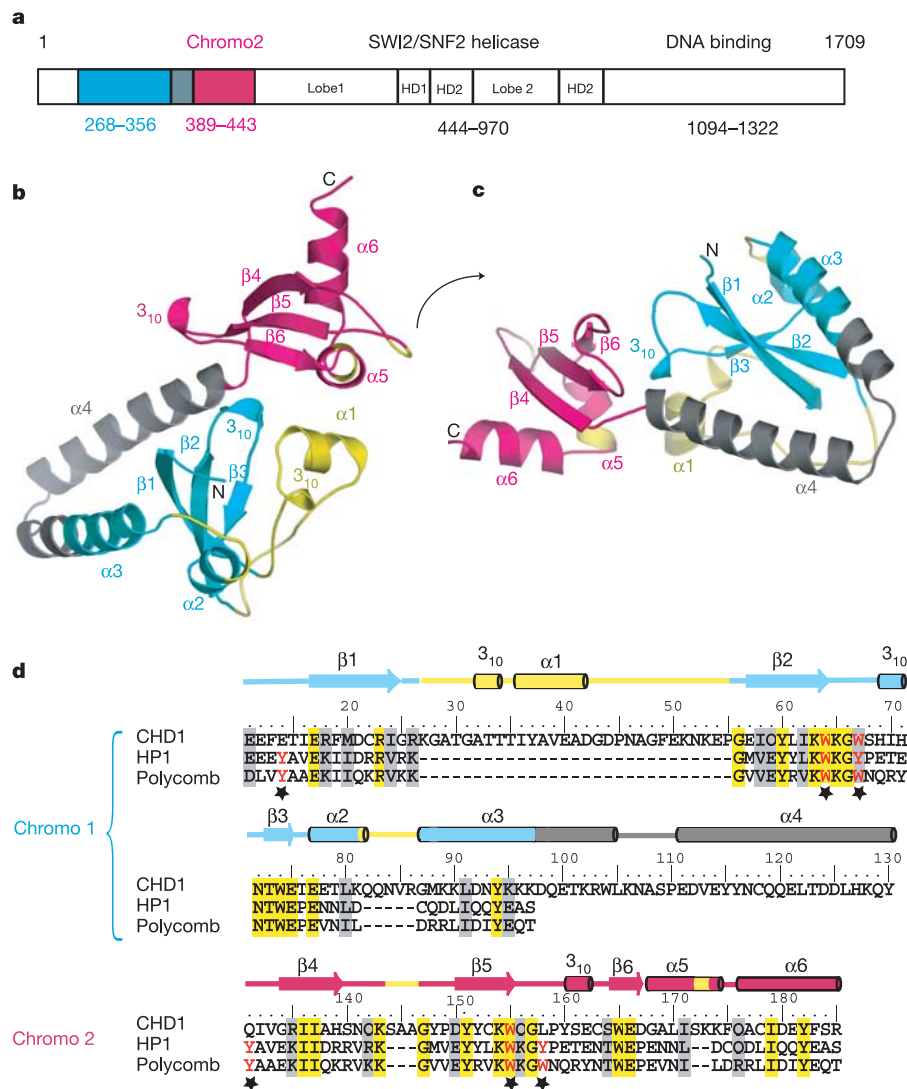


Figure 1 | Structure of human CHD1 double chromodomains. **a**, The conserved domains within human CHD1 sequence. The SWI2/SNF2 region is related to that of the Rad54 protein²⁷. **b**, **c**, Crystal structure of the tandem chromodomains in two views. Cyan, chromodomain 1; grey, linker; pink, chromodomain 2; yellow corresponds to inserted sequences compared with

those of HP1 and Polycomb. **d**, Structure-based sequence alignment of CHD1 individual chromodomains with HP1 and Polycomb. The secondary structure elements above the sequence correspond to those in CHD1. The black stars indicate aromatic residues that contribute to the cage in HP1 and Polycomb.

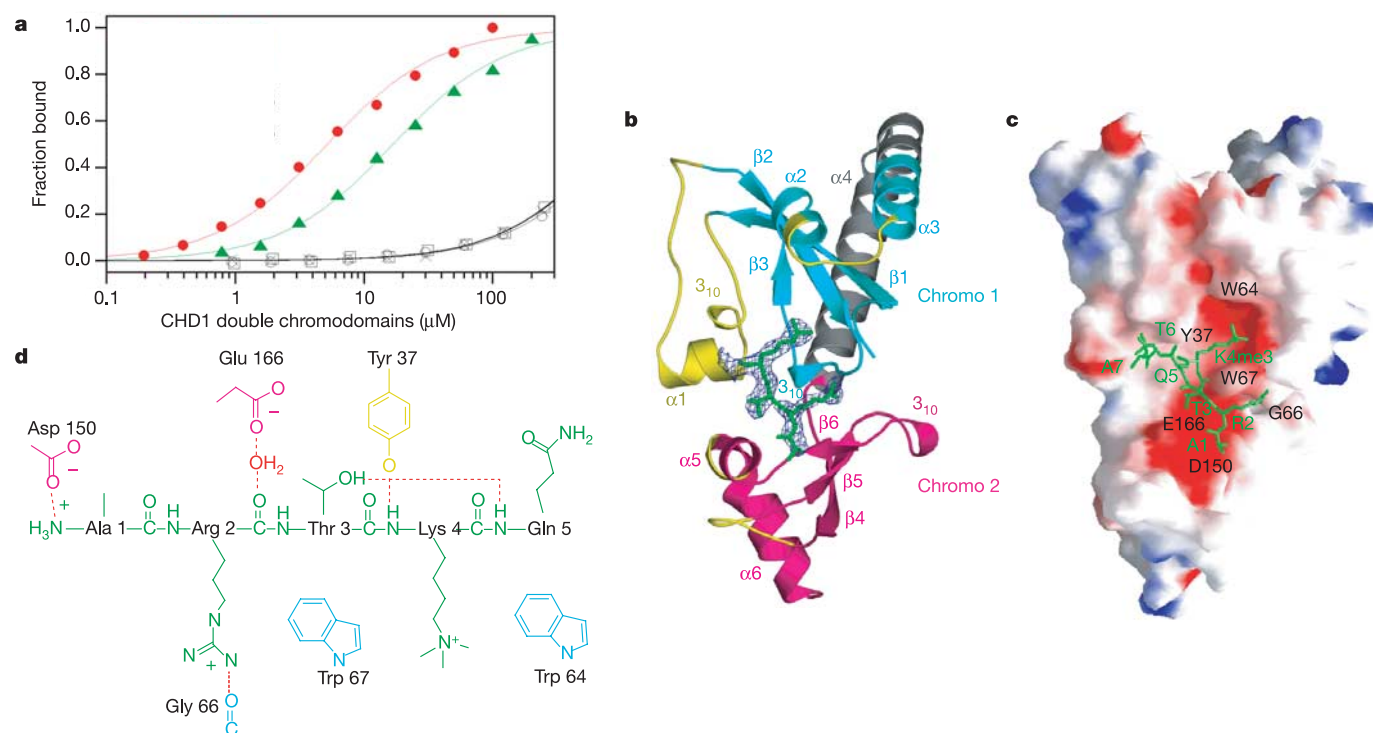


Figure 2 | Peptide selectivity of the human CHD1 double chromodomains. **a**, Fluorescence polarization binding assays (Supplementary Table S2). Red filled circles, H3K4me3; green triangles, H3K4me1; black crosses, H3K36me3; black squares, H3K79me3; black open circles, unmodified H3 tail. **b**, The structure of H3K4me3 bound to CHD1. Blue, electron density from an $|F_o - F_c|$ omit map contoured at 3σ where the H3 peptide was

omitted for the map calculation. H3 peptide is traced in green. **c**, Surface electrostatic potential showing the peptide (green)-binding groove on CHD1 (ref. 28), in the same view as in **b**. **d**, Determinants of peptide binding as seen in the structures involving H3K4me3 and H3K4me1 peptides.

dephosphorylation seems to allow CHD1 to be reincorporated into chromatin during telophase¹⁹. Lysine 4 methylation and threonine 3 phosphorylation therefore seem to act as a binary switch²⁰ for the binding of CHD1.

To further understand the effect of phosphorylation, we analysed the binding of CHD1 chromodomains to the H3 tail peptide with simultaneous phosphorylation at threonine 3 and trimethylation at lysine 4. Figure 3d shows the binding to be 25-fold weaker than the binding to methyllysine 4 alone. To understand the basis for the weaker binding we solved the structure of the complex with the doubly modified peptide (Fig. 3a). The structure shows a loss of an intrapeptide hydrogen bond between the side chain hydroxyl of threonine 3 and the backbone amide of glutamine 5 when the former is phosphorylated (Fig. 2d). The high crystallographic *B* factors in the peptide in the H3K4me3T3ph complex (Supplementary Table S1) are probably a result of the low occupancy of the peptide due to low affinity, or the reduced stability of the bound peptide conformation. More distal modifications with respect to methyllysine 4, namely acetylation or methylation of lysine 9, phosphorylation of serine 10 or acetylation of lysine 14, do not perturb the *K_d* of binding to the lysine 4-methylated H3 tail (Supplementary Table S2).

The single chromodomains of HP1 and Polycomb proteins also bind to the histone H3 tail but to two other regions. HP1 binds to methyllysine 9, using residues 5–10 for specificity, whereas Polycomb binds to methyllysine 27 and using residues 20–28 for specificity. In each case the H3 tail inserts as a β -strand to complete the overall β -sandwich fold of the chromodomain in HP1 and Polycomb (Fig. 4). Figure 1d shows that CHD1 chromodomain 1 contains two significant insertions in its protein sequence relative to HP1 and Polycomb. One insertion, residues 27–55, has a clear functional role, helping to establish the peptide-binding site through tyrosine 37 (Figs 1d and 2b, d).

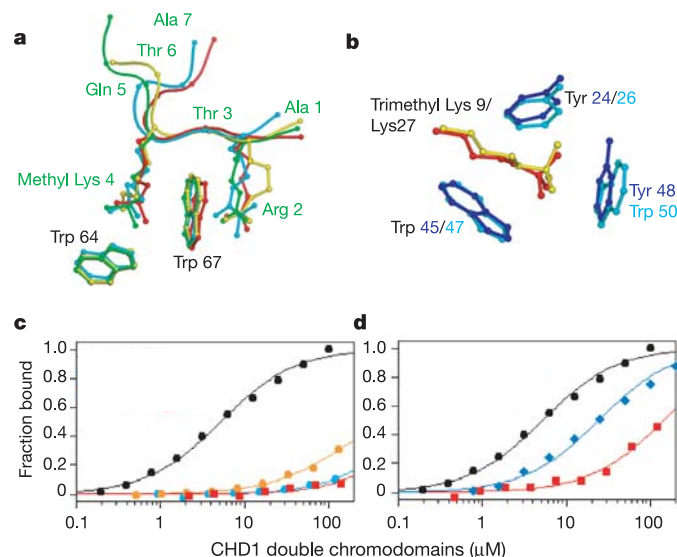


Figure 3 | Methyllysine binding by human CHD1. **a**, The bound-peptide structure of H3K4me3 (green), H3K4me1 (yellow), H3K4me3R2me2a (cyan) and H3K4me3T3ph (red). **b**, The aromatic cages from HP1 (blue) and Polycomb (cyan) with methyllysines 9 and 27. **c**, Fluorescence polarization peptide binding assays using human or yeast CHD1. Using H3K4me3 peptide, human CHD1 (black) binds with a *K_d* of 5 μM, W67L-human CHD1 (cyan) does not bind, W64L-human CHD1 (yellow) binds with a *K_d* of 290 μM, and budding yeast CHD1 (red) does not bind. W64L and W67L-human CHD1 do not allow binding to H3K4me1 or the unmodified peptide. **d**, Effect of adjacent peptide modifications. Black, H3K4me3; cyan, H3K4me3R2me2a; red, H3K4me3T3ph.

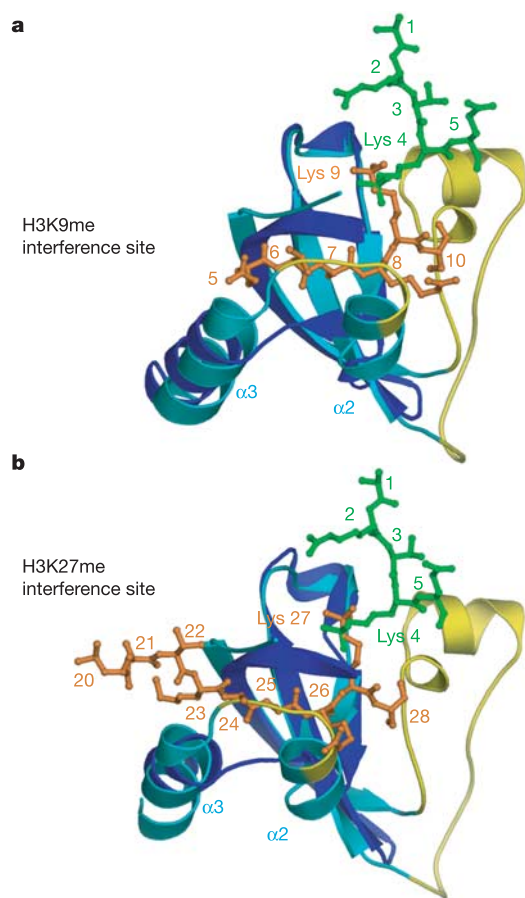


Figure 4 | Comparison of CHD1 with HP1 and Polycomb. **a**, HP1 chromodomain (blue) in complex with H3K9me3 (brown) superimposed on CHD1 chromodomain 1 (cyan and yellow) in complex with H3K4me3 (green). **b**, Polycomb chromodomain (blue) in complex with H3K27me3 (brown) superimposed on CHD1 chromodomain 1 (cyan and yellow) in complex with H3K4me3 (green).

There is also functional significance for the second insertion (residues 82–86) in chromodomain 1 (Fig. 1d). This insert, between helices $\alpha 2$ and $\alpha 3$, occupies the site used for the recognition of the H3 tails by the HP1 and Polycomb proteins (Fig. 4). Thus, this loop blocks and interferes with the expected site of peptide interaction on chromodomain 1 of CHD1. Chromodomain 2 of CHD1 is also disabled from binding the H3 tail with the canonical interactions seen in HP1 and Polycomb, because of the availability of only one conserved aromatic residue (tryptophan 155) for binding to methyl-lysine (Fig. 1d). Because neither chromodomain alone is able to bind the peptide, the two chromodomains instead cooperate to create the recognition site.

Other conserved modules have been implicated in the recognition of post-translationally modified chromatin²¹, and appear in certain proteins as tandem repeats. The tandem bromodomains of TAFII250 protein bind selectively to multiply acetylated histone H4 peptides. It has been suggested that these two bromodomains form a side-by-side surface with two independent acetyllysine-binding pockets that are ideal for the recognition of one diacetylated histone H4 tail²². The 53BP1 tandem tudor domains are also implicated in binding to the core of histone H3 when lysine 79 is methylated, and it has been suggested that residues that participate in the recognition of H3 lie at the interface of the two tudor domains²³. Further structure analyses involving target peptides are needed to address the significances of these double arrangements.

METHODS

Protein preparation. All expression constructs contained an N-terminal His₆ tag, were cloned into the *Bam*HI/*Nde*I sites of the pET11a vector, expressed in BL21(DE3) *Escherichia coli* (Novagen) and purified by Ni²⁺-affinity chromatography (Qiagen). Point mutations were prepared with QuikChange (Stratagene). For human CHD1, the chromodomains 1 + 2 construct corresponds to residues 268–443, and the chromodomain 1 construct (with a point mutation C280S) corresponds to residues 268–362. For yeast CHD1, the chromodomains 1 + 2 construct corresponds to residues 174–339, and the chromodomain 2 construct corresponds to residues 282–339.

Binding assays. For fluorescence polarization, 100 nM fluorescein-labelled peptide (prepared as described previously²⁴) was used in 50 mM sodium phosphate pH 8.0, 25 mM NaCl, 5 mM tris(2-carboxy-ethyl) phosphine (TCEP). For isothermal titration calorimetry (ITC) both protein and peptide were dialysed into 20 mM bis-tris propane (BTP) pH 8.0, 25 mM NaCl and 10 mM BME. Injections (10 μ l) of H3K4me3 peptide at 850 μ M were titrated into protein at 70 μ M concentration, with the use of a published protocol²⁴.

X-ray crystallography. Crystals were prepared at 10 °C with protein samples (13.75 mg ml⁻¹) in 20 mM BTP pH 8.0, 25 mM NaCl and 10 mM TCEP. For peptide complexes, dry peptide was added to 2.5 mM. Crystals of both free and complex grew from 6% PEG3350, 100 mM HEPES pH 8.0, and were cryo-protected in 8% PEG3350, 100 mM HEPES pH 8.0 and 35% ethylene glycol. There were 2½ molecules per asymmetric unit cell (two intact CHD1 polypeptides plus a truncation product consisting of chromodomain 1). The selenomethionyl derivative was prepared by expression in B834(DE3) *E. coli* grown in minimal media supplemented with Seleno-Met (Sigma). The structures of the free protein and complexes with peptides were determined by using the model of the H3K4me3T3ph complex solved by multiwavelength anomalous diffraction as the search model for molecular replacement (Supplementary Table S1). The program CNS was used for molecular replacement, and each model was improved by rigid body and simulated annealing refinement²⁵. Figures were prepared with PyMOL²⁶.

Received 12 July; accepted 4 October 2005.

1. Brehm, A., Tufteland, K. R., Aasland, R. & Becker, P. B. The many colours of chromodomains. *BioEssays* **26**, 133–140 (2004).
2. Tajul-Arifin, K., Teasdale, R., Ravasi, T., Hume, D. A. & Mattick, J. S. Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Res.* **13**, 1416–1429 (2003).
3. Lusser, A., Urwin, D. L. & Kadonaga, J. T. Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nature Struct. Mol. Biol.* **12**, 160–166 (2005).
4. Bouazoune, K. *et al.* The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J.* **21**, 2430–2440 (2002).
5. Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P. & Collins, F. S. Characterization of the CHD family of proteins. *Proc. Natl Acad. Sci. USA* **94**, 11472–11477 (1997).
6. Stokes, D. G., Tartof, K. D. & Perry, R. P. CHD1 is concentrated in interbands and puffed regions of *Drosophila* polytene chromosomes. *Proc. Natl Acad. Sci. USA* **93**, 7137–7142 (1996).
7. Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E. & Schreiber, S. L. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395**, 917–921 (1998).
8. Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279–289 (1998).
9. Schneider, R. *et al.* Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biol.* **6**, 73–77 (2004).
10. Jacobs, S. A. & Khorasanizadeh, S. Structure of the HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
11. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
12. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
13. Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
14. Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434–438 (2005).
15. Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol. Cell* **12**, 1325–1332 (2003).
16. Schurter, B. T. *et al.* Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry* **40**, 5747–5756 (2001).

17. Kouskouti, A. & Talianidis, I. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *EMBO J.* **24**, 347–357 (2005).
18. Dai, J., Sultan, S., Taylor, S. S. & Higgins, J. M. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes Dev.* **19**, 472–488 (2005).
19. Stokes, D. G. & Perry, R. P. DNA-binding and chromatin localization properties of CHD1. *Mol. Cell. Biol.* **15**, 2745–2753 (1995).
20. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
21. Khorasanizadeh, S. The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259–272 (2004).
22. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAF_{II}250 double bromodomain module. *Science* **288**, 1422–1425 (2000).
23. Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411 (2004).
24. Jacobs, S. A., Fischle, W. & Khorasanizadeh, S. Assays for the determination of structure and dynamics of the interaction of the chromodomain with histone peptides. *Methods Enzymol.* **376**, 131–148 (2004).
25. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
26. DeLano, W. L. *PyMOL User's Guide* (DeLano Scientific, San Carlos, California, 2004).
27. Thoma, N. H. *et al.* Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nature Struct. Mol. Biol.* **12**, 350–356 (2005).
28. Nicholls, A. *GRASP: Graphical Representation and Analysis of Surface Properties* (Columbia University, New York, 1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Zimmerman for assistance with diffraction data collection. This work was supported by grants from the National Institutes of Health (to S.K.).

Author Contributions J.F.F. and L-Z.M. contributed equally to this work.

Author Information The atomic coordinates have been deposited in the Protein Data Bank with the accession numbers 2B2Y, 2B2W, 2B2V, 2B2U and 2B2T. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.K. (khorasan@virginia.edu) or F.R. (fr9c@virginia.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Meeting pay-offs

Going to a meeting means spending time away from the lab, often involves expensive travel costs and, occasionally, can see scientists having to overcome problematic security or visa issues. But is this hassle really worthwhile?

The scientific community in general seems to think so: despite the advent of e-mail and video conferencing researchers still flock to meetings around the world. But until recently hard data showing the value of such events have been hard to find. A recent survey by Keystone Symposia, a non-profit meetings organizer in Silverthorne, Colorado, which admittedly looked only at its own meetings, has gone some way to address this lack of information. It suggests that scientists going to meetings can save six weeks of research time and US\$6,000 in funding (see *Nature* **438**, 264-265; 2005).

The survey is useful because it might help researchers justify the expense of a proposed trip. The data could be especially helpful in the present climate of tight budgets and tighter restrictions. For example, researchers working for the US National Institutes of Health now

have to make a very strong case to attend conferences.

But one thing the survey doesn't look at is how scientists can make the most of their time at a conference. The benefits of meetings — building fresh collaborations, gaining new insight — do not arise automatically; they require work. Few scientists receive formal training in setting goals and drawing up strategies for maximizing their time at a conference. Earlier in the year, *Naturejobs* listed a few ideas that we hope will make meetings more productive (see *Nature* **436**, 1060-1061; 2005).

Perhaps one New Year's resolution should be to apply these strategies to 2006 meetings, including those listed in *Nature's* annual events directory. After all, coming home from a conference with concrete accomplishments is the best justification for going in the first place.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

The Quantum before Christmas

In search of the sanity clause.

Henry Gee

Dear Children,

Thank you for your kind letter. A DVD of *The Polar Express* and the latest Gorillaz album should present no problems. We shall see what can be done about getting your father to take part in *The X-Factor*, especially given the sonorous quality of his bath-time rendition of *Unchained Melody*, although there are, as you will appreciate, issues related to surface-area-to-volume ratio. A viable velociraptor might be more problematic, especially as you already have two cats, who might object.

As to your other inquiries of a more personal nature — despite the fact that these are somewhat off-topic, these are questions that many people such as yourself ask at this time of year, and answers can (of course) be provided.

First: whether I exist. Ah, existence. This is one of those things concerning which everyone ties themselves up in horrible topological incongruities — everyone, that is, except me, if indeed it is 'me' to whom I am (self-referentially) addressing this comment. For what is 'existence' but the shadow of an impression of collapsed wave-functions? Long experience suggests that individuality is related to sentience, and with that, pain and the fear of death. At this point all inquiries can be referred to Keats' *Ode To A Nightingale* (ask your father to read this to you when you are older). Anyway, if there is no 'me', how can 'I' exist such that 'my' absence would mean anything to...er... 'me'? And what do 'I' get for indulging in such existential crises? Here's what — *bupkes!* Or as the Good Book says, if there is no self, whose arthritis is this? In ignorance lies happy immortality.

In any case, if the Gorillaz CD is to be secured, let alone the velociraptor, there's no time to waste, not that 'time', like individuality, is a topic worth wasting 'time' to discuss. So let's get on with it. Oh all right, if you insist. The great thing about there being no 'me' is that people can ascribe all kinds of properties to the entity with which I am congruent, without this entity,

whoever it is, minding in the least. So let's let you into a secret — the hat and the boots are real (not that 'reality'...oh, never mind), but the reindeer are fairly recent inventions. Simple aerodynamics dictates that reindeer of the size of the conventional ungulate cannot get airborne. It's all a matter of Reynolds numbers, apparently.

Your second point — whether it is possible to visit all the good children of the world in just one night — the answer is an emphatic 'yes'. The reason is related to your first question, for the facility to achieve this involves necessary compromises in the fields of existence, individuality, time and reality. No need to discuss hypersonic shock waves and the inevitable problems of squeezing down non-existent chimneys in centrally heated houses, for it is possible to be in an arbitrarily large number of places simultaneously, because, as your father has no doubt explained, 'I' am a macroscopic quantum object. Please don't feel badly that you didn't believe your father when he explained this to you, for more experienced minds than yours have grappled with this selfsame concept. In the words of Newton: "Be here now, be someplace else later: is that so complicated?" to which can be added Einstein's corollary: "Wherever you go, there you are. Your luggage is another story."

So much is clear, but the Universe is a harsh mistress, and exacts a price for such facility. Or, rather, two. The first is that one must remain rather chilly — for at temperatures any greater than the achingly frigid, one loses even the capacity to discuss such concepts as individuality, whether or not such things apply to one's own state. You will no doubt have wondered — or if you hadn't, you should have — why you addressed your letter to the North Pole in the middle of winter, and not (for example) Florida, a place with manifest attractions to one such as 'myself' (in my traditional jolly-white-haired-grandfather avatar), and in which individuality is, in any case, neither here nor there. Ah! Even the existentially dissipated can dream. But the mince pies will be very much appreciated. And the sherry.

The second cost is loneliness. Aha, you might say, how can one who lacks individuality suffer from such a malady? 'I' don't know the answer to that either, except that after several eternities, the lack of decent conversation rather gets one down, which is why answering letters such as yours is so therapeutic. But it goes with the job — it is not something that can be shared, because it must be carried out in absolute secrecy. The consequences were anyone to blunder in on 'my' operations would be utterly disastrous, by virtue of Heisenberg's uncertainty principle. Not to put too fine a point on it, my wave-function would collapse, a sensation which

(despite all the caveats above) is probably best not experienced, and would in all probability cause a great deal of inconvenience to everyone. From this it should be clear why you must be tucked up in bed and sound asleep well before midnight on 24 December.

With the compliments of the season,
Sincerely,
pp Santa.

P.S. Don't tell your father you received this. I have already bribed the postman.

Henry Gee is a senior editor of *Nature* and wears woolly socks, size 11...just what I always wanted, no, you shouldn't have.



JACEY